

Приветствие

Этот курс позволит вам погрузиться в удивительный мир квантового машинного обучения!

Почему именно этот курс?

Наш курс отличается от других курсов по квантовым вычислениям:

- он адаптивный и содержит лекции разных уровней сложности и глубины;
- он практический, а все объяснения подкрепляются кодом;
- он про реальные методы, которые будут актуальны ближайшие 10-15 лет.

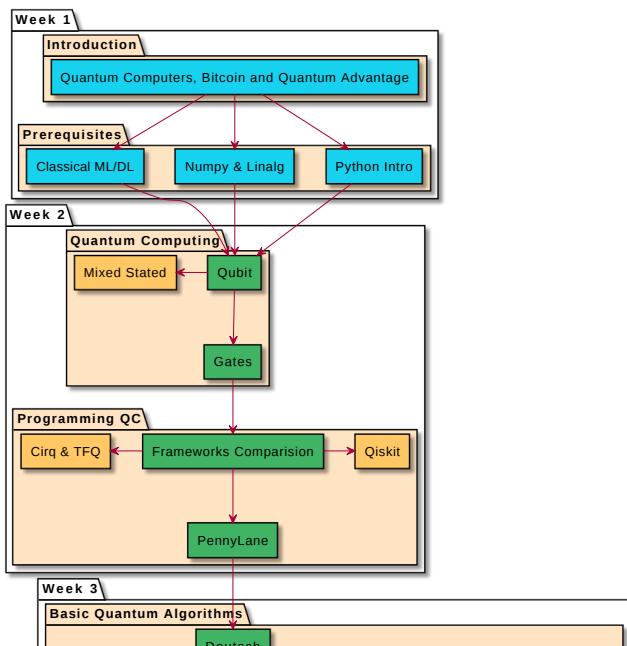
Основатель курса [Семён Синченко](#):

Я очень люблю квантовые технологии и являюсь их большим энтузиастом. При этом моя основная работа связана с машинным обучением, поэтому идею сделать курс по QML я вынашивал очень давно. В сообществе ODS мне удалось найти единомышленников, и вместе мы подготовили для вас набор лекций и домашних заданий, которые проведут вас через удивительный мир кубитов, гейтов и квантовых нейросетей. Мы очень хотели сделать курс одновременно простым и понятным, но также строгим и глубоким с технической точки зрения и, надеюсь, у нас это получилось. В основном мы будем обсуждать темы, которые будут перспективны в ближайшие годы, а также максимально ориентироваться на практической стороне вопроса. Я лично верю, что квантовые технологии по мере их развития произведут не меньшую революцию, чем, например, распространение вычислений на видеокартах, а также позволят по-другому подойти к задачам, которые мы раньше не умели решать . Надеюсь, что после прохождения этого курса поверите и вы! В любом случае, я уверен, что тема QML и квантовых вычислений как минимум стоит того, чтобы с ней познакомиться. Надеюсь, что вам будет интересно!

Как устроен курс?

Наш курс разделен на логические блоки, каждый из которых содержит лекции разных уровней сложности:

- **ГОЛУБОЙ** – вводные лекции;
- **ЗЕЛЕНЫЙ** – лекции “основного” блока курса;
- **ЖЕЛТЫЙ** – лекции, глубже раскрывающие темы блоков;
- **КРАСНЫЙ** – лекции про физику и математику, которая стоит за всем этим;
- **БЕЛЫЙ** – факультативные лекции.



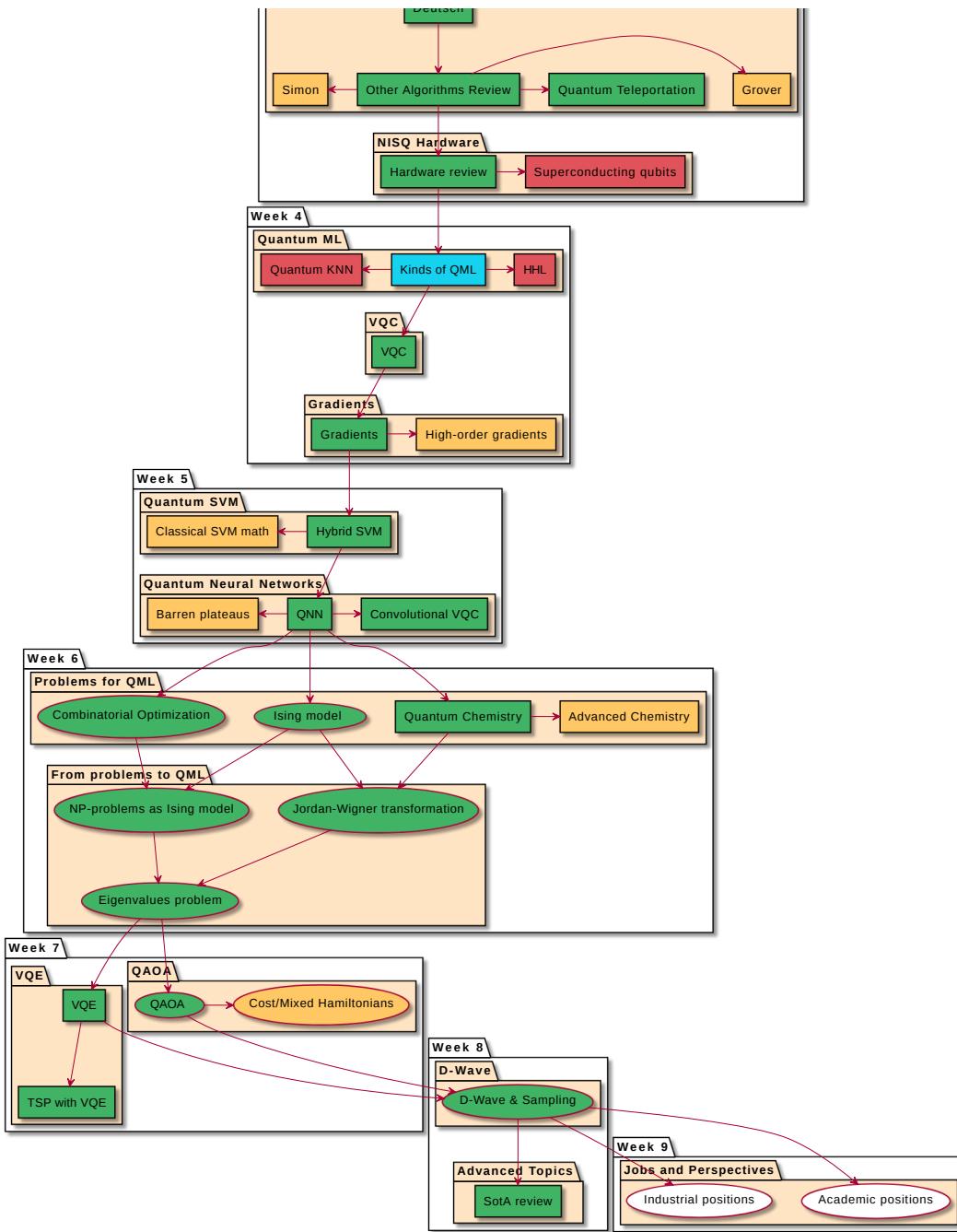


Fig. 1 Программа курса

Как проходить этот курс?

Рекомендуется проходить курс в том порядке, который обозначен на схеме. У курса есть [пререквизиты](#) – программирование на Python, основы линейной алгебры и машинного обучения, советуем с ними ознакомиться и (опционально) проверить себя с помощью [входного теста](#).

Курс обсуждается в [Slack ODS.ai](#) в канале #quantum_computing.

Как можно помочь?

Помочь можно словом и делом. Если знаете людей, интересующихся квантовыми вычислениями – замолвите словечко. Курс разрабатывается сообществом OpenDataScience, такими же «большими», как и организаторы ([@sem](#) [@yorko](#) [@vtrohimenko](#)). Если горите желанием в чем-то помочь, то создавайте Issue, а лучше сразу Pull Request в [репозитории курса \(Contributing guide\)](#).

Желаем успехов!

О квантовых компьютерах, биткоине и превосходстве

Описание лекции

Эта лекция не несет в себе образовательного смысла, а лишь пытается ответить на вопросы, которые обычно возникают у тех, кто впервые сталкивается с темой квантовых вычислений. А именно:

- что это за вычисления такие?
- зачем вообще это все нужно?
- и когда взломают биткоин?
- что за превосходство, о котором все говорят?

Что это вообще за компьютеры такие?

Количественная эволюция компьютеров

Сегодня классические компьютеры, построенные на идеях Тьюринга, фон Неймана и Шокли, стали неотъемлемой частью нашей жизни. Все мы привыкли к тому, что с каждым годом наши компьютеры становятся все мощнее и мощнее. И то, что сегодня является бюджетным ноутбуком 15 лет назад было аналогом суперкомпьютера!



Fig. 2 Иллюстрация закона Мура — рост числа транзисторов с 1970-х

Так называемый закон Мура, сформулированный Гордоном Муром еще в конце 60-х годов, говорит нам о том, что число транзисторов на кристалле интегральной схемы будет удваиваться каждые два года. И этот закон стабильно выполняется.

Качественная эволюция компьютеров

Но, к сожалению, по сугубо физическим причинам, выполнение закона Мура не может длиться вечно — рано или поздно, но прямое увеличение количества транзисторов станет невозможным. Поэтому сегодня все чаще можно услышать слова о том, что современным компьютеры ждет качественная революция. Кто-то говорит о переходе на новые материалы для изготовления транзисторов. Кто-то говорит о создании транзисторов на новых принципах, например, об оптических компьютерах. Но часто можно услышать слова о том, что следующим революционным прорывом станет создание квантовых компьютеров. О них мы и будем говорить.

Идея о квантовом компьютере

Сегодня существует несколько версий о том, кто же первым высказал идею о квантовом компьютере. Как это часто бывает, сразу несколько ученых одновременно и независимо пришли к одной и той же идее. Одним из таких ученых был Ричард Фейнман.



Fig. 3 Ричард Фейнман, 1918-1988

В 1981-м году, когда шло очень активное развитие одновременно классических компьютеров и вычислительной квантовой механики, он высказал идею о том, что для решения задач квантовой физики нам нужен квантовый компьютер.

Что это за компьютер такой?

Этот вопрос на самом деле крайне сложный и именно ему будет посвящена первая половина нашего курса. Кажется странным, если вопрос, которому будет посвящено несколько полноценных лекций с формулами можно было бы раскрыть в одном абзаце. Было бы ошибкой пытаться сказать, что обычные компьютеры работают на законах классической физики, а квантовые на основе квантовой механики — ведь нормально объяснить работу транзистора можно лишь с привлечением уровня Ферми и прочих сложных штук из квантовой физики. Также неправильно было бы говорить о том, что в отличии от классических компьютеров, где есть лишь $|0\rangle$ и $|1\rangle$ в квантовых есть все состояния сразу. Ведь ничего не мешает сделать так называемую вероятностную машину Тьюринга, другими словами, классический компьютер, который оперирует многими состояниями сразу. Особенно не хочется сразу сыпать на читателя кучу непонятных терминов, типа квантовой суперпозиции, кубита или запутанности, ведь для тех, кто не знает что такое квантовые компьютеры эти термины, вероятнее всего, тоже ничего не дадут. Для начала, давайте просто условимся, что квантовые компьютеры это, в отличии от фотонных, графеновых, или других перспективных "новых" компьютеров, не только использование новых материалов или технической базы, а еще и использование новой, отличной от заложенной Тьюрингом концепции вычислений, представления и обработки информации.

А зачем это вообще нужно?

Факторизация больших чисел

Мне кажется, что именно открытие алгоритма Шора для эффективного решения задачи факторизации послужило наибольшим толчком в популяризации квантовых вычислений. Именно после этого большое число специалистов устремилось в эту область, военные и корпорации начали вкладывать деньги, а журналисты стали писать о будущем крахе банковских платежей и вообще всего мира. Вероятно, алгоритм Шора является самым раскрученным квантовым алгоритмом.

Дело в том, что большая часть всей современной криптографии держится на одном простом предположении о невозможности эффективно решать задачу факторизации больших чисел. Ну или по простому, если у нас есть число, которое является произведением двух относительно больших простых чисел, то мы будем до бесконечности искать эти числа и скорее всего так и не найдем их. Но это для классического компьютера. А вот для квантового компьютера в 1994-м Питером Шором был предложен алгоритм, который решает эту задачу эффективно, за относительно короткий промежуток времени.



Fig. 4 Питер Шор, тот, кто переполошил весь мир своим алгоритмом

Именно этот алгоритм в будущем взломает биткоин и обрушит банковскую систему, но не все так плохо – развитие квантовых компьютеров подтолкнуло ученых в области криптографии к созданию новых, так называемых *пост-квантовых* алгоритмов шифрования, которые нельзя взломать за разумное время и на классическом, и на квантовом компьютере.

Комбинаторные и NP-трудные задачи

Помимо уже озвученных проблем с выполнением закона Мура, есть также и другая проблема. А именно то, что существуют задачи, которые скорее всего никогда не получится эффективно решать на классическом компьютере Тьюринга. Даже на фотонном или графеновом. Хороший пример это задача о рюкзаке. Когда у нас есть рюкзак ограниченного объема, а также есть много предметов разного веса и стоимости. И нам надо наполнить наш рюкзак так, чтобы предметы внутри него имели максимальную суммарную стоимость. Задача кажется легкой, но она относится к так называемым \(NP\)-полным задачам. Такие задачи, например, в случае большого рюкзака и набора предметов, невозможно точно решить за разумное время. Да и вообще их решить с приемлемой точностью, пусть даже *приближенно* это сегодня большая проблема!

Note

Здесь я не зря написал “скорее всего”. Дело в том, что этот вопрос является одним из так называемых вопросов тысячелетия. Так, для известной задачи о наполнении рюкзака, мы не знаем сегодня эффективного алгоритма решения на классическом компьютере. Но мы также и не можем пока доказать, что такого алгоритма не существует. Ну то есть скорее всего такого алгоритма и правда не существует, а также скорее всего $\text{P} \neq \text{NP}$, но доказать это пока ни у кого не вышло. Но это скорее лирическое отступление.

Так вот, дело в том, что для квантовых компьютеров уже сегодня известны алгоритмы, которые позволяют потенциально эффективно, пусть и *приближенно* решать такие задачи на квантовом компьютере. Это задача коммивояжера, задача о рюкзаке, задача кластеризации графа и много других задач комбинаторной оптимизации. В нашем курсе будет целый блок, посвященный таким квантовым алгоритмам как *Variational Quantum Eigensolver* и *Quantum Approximate Optimization Algorithm*.



Fig. 5 Визуализация решения задачи коммивояжера – кратчайший путь, чтобы обехать 12 немецких городов – очень важная задача современной логистики

Симуляция квантовой механики

Это то, ради чего Фейнман предложил создать квантовые компьютеры. Это отдельная большая тема, где много квантовой механики. Ей будет посвящено сразу несколько отдельных лекций нашего курса. Но попробуем объяснить в двух словах, не вдаваясь в детали.

Дело в том, что задачи квантовой механики не получается решать аналитически. Казалось бы, в чем проблема, законы Ньютона уже для трех тел тоже аналитически не решаются, но это не мешает нам летать в космос, ведь такую задачу можно решить **численно**. Но тут приходит вторая проблема, а именно, что явно интегрировать уравнение Шредингера по времени, или, по простому, решать квантовую механику **численно** тоже вычислительно почти невозможно более чем для двух частиц.



Fig. 6 Эрвин Шредингер, 1887-1961, создатель знаменитого уравнения и мема про кота

Казалось бы, что нам с этого. Ведь квантовая механика это удел теоретиков. Но вот проблема, квантовая механика лежит в основе квантовой химии, а та, в свою очередь, лежит в основе вообще всей химии и таких ее прикладных направлений, как создание новых лекарств, разработка новых аккумуляторов для автомобилей Tesla и многое другое. И сегодня мы вынуждены использовать лишь очень приближенные решения и концепции, точности которых часто не хватает.

Квантовые компьютеры в этом случае могут сделать реальный прорыв. Ведь в силу своей физической природы, квантовый компьютер идеально подходит для симуляции квантовой механики, а значит и решения столь важных сегодня задач из области разработки лекарств и дизайна новых материалов.

Машинное обучение и искусственный интеллект

За последние 10-15 лет машинное обучение достигло поистине небывалых высот в своем развитии. Многие задачи, решение которых силами компьютера, раньше казалось невозможным сегодня успешно решаются при помощи машинного обучения. Примеры таких задач это, например, игра в Go, различение пород чихуахуа по фотографии, распознавание лиц в видеопотоке, составление относительно осмысленных текстов и генерация картин в стиле Пикассо из простых фотографий. Но оно все еще очень далеко от возможностей человеческого мозга. Так, наиболее масштабные искусственные нейронные сети, по примерным оценкам, имеют сегодня размер, эквивалентный 15 миллионам нейронов, в то время как человеческий мозг имеет порядка 85 миллиардов! Вызывает вопросы также и скорость обучения современных нейронных сетей. Так, самые большие языковые модели сегодня обучаются неделями на кластерах из тысяч видеокарт, в то время как человек с его относительно скромными вычислительными возможностями учится говорить всего 2-3 года.

И тут тоже на помощь могут прийти квантовые компьютеры. В данном случае, квантовые аналоги нейронных сетей, а также их комбинации с классическими нейронными сетями уже сегодня показывают впечатляющие результаты. Так, есть работы, где показано, что 4 квантовых нейрона по своей выразительности эквивалентны классической искусственной нейронной сети с ~ 250 нейронами!

Именно квантовому машинному обучению, а также способам его применения и будет посвящена большая часть нашего курса. Мы постараемся рассмотреть все вопросы по этой теме, начиная от теории того, как можно строить квантовые алгоритмы машинного обучения и заканчивая тем, как их можно запрограммировать на современных языках квантового программирования. Если эта тема вам интересна, то этот курс точно для вас!

Ну и когда взломают биткоин?

Наверное это один из главных вопросов, которые возникают при чтении подобных статей. И ответим сразу: взломают нескоро, времени еще много, 10 лет точно есть.



Fig. 7 Биткоин, как и многие другие электронные средства вынуждены будут перейти на пост-квантовую криптографию

Сколько нужно кубитов под разные задачи?

Наверное сразу стоит оценить тот размер, который квантовый компьютер должен иметь для эффективного решения описанных выше задач. Примерно цифры такие:

- Алгоритм Шора и взлом современной криптографии (включая биткоин): $\sim 20 \cdot 10^6$ (20 миллионов) кубит
- Задачи оптимизации: $\sim 100 \cdot 10^3$ (100 тысяч) кубит
- Первые полезные задачи в квантовой химии: $\sim 1 \cdot 10^3$ (1 тысяча) кубит
- Квантовое машинное обучение: $\sim 100-500$ кубит

Это кстати одна из причин, почему наш курс посвящен по большей части именно квантовому машинному обучению.

Логические vs Физические кубиты

Есть еще такая проблема, что вся квантовая механика вероятностная. А еще, что квантовые компьютеры работают в области микромира и очень чувствительны к любым шумам извне. Это ведет к совершенно недопустимому уровню ошибок в вычислениях и их низкой детерминированности. Например, сегодня хорошим

уровнем точности для квантовых компьютеров является 99% на одну операцию. Но ведь каждый алгоритм включает в себя сотни или даже тысячи операций! И тогда уровень ошибок становится совсем печальным.

Но есть и хорошие новости. Сегодня существует очень много классных алгоритмов коррекции ошибок, которые позволяют используя несколько физических кубит с высоким уровнем ошибок создать один логический кубит, имеющий очень низкий уровень ошибок. То есть программист будет писать код, который производит операции над одним кубитом, а на физическом уровне это будет операция над несколькими кубитами. В общем вопрос вполне решаемый. Вот только для создания одного качественного логического кубита может потребоваться до тысячи физических кубит! А те оценки, которые мы привели выше, они как раз про логические кубиты, то есть кубиты с очень высокой точностью операций на уровне классических компьютеров.

Сколько кубит есть сегодня?

Скажем сразу, сегодня уже существуют квантовые компьютеры. Вот только все производители, когда пишут о новом рекорде, имеют в виду чаще всего именно физические кубиты.

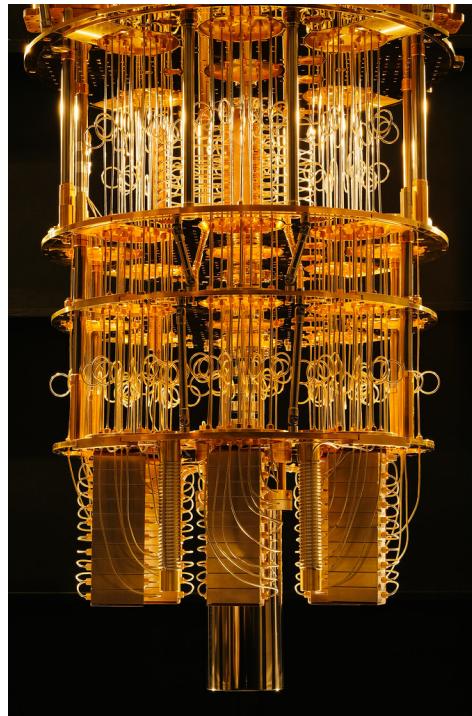


Fig. 8 Квантовый компьютер компании IBM выглядит примерно так

Есть квантовые компьютеры с разной архитектурой. Одни имеют больше кубит, но и более высокий уровень ошибок. Другие имеют низкий уровень ошибок, но их трудно масштабировать. Тема квантового железа в нашем курсе будет посвящен отдельный блок из нескольких лекций. Но если кратко, то можно назвать примерно такие цифры:

- рекорд в относительно легко масштабируемых, но шумных квантовых компьютерах это $\backslash(\backslashsim 55)$ кубит
- рекорд в относительно точных, но медленных и плохо масштабируемых компьютерах это $\backslash(\backslashsim 20)$ кубит
- рекорд в точных и масштабируемых, но очень трудно программируемых компьютерах это $\backslash(\backslashsim 25)$ кубит

Note

Тут мы имеем ввиду соответственно:

- сверхпроводящие кубиты, которые сегодня проще всего масштабировать
- ионы в ловушках, которые имеют одну из самых высоких точностей
- фотоны, которые вроде всем хороши, кроме того, что на них программирование это юстировка линз и лазеров на оптическом столе

Стоит добавить, что рекорд в точных и масштабируемых, а также программируемых (топологических) кубитах сегодня это ровно 2 кубита. Серьезно, взаимодействие двух логических кубит было опубликовано в *Nature* в этом году.

Какие планы имеют ведущие игроки на этом рынке?

Казалось бы, с такими масштабами биткоину боятся нечего, да и в целом область выглядит не самой перспективной. Но есть один нюанс. Все крупные игроки на рынке создания квантовых компьютеров (*Google Quantum*, *IBM Quantum*, *IonQ*, *Xanadu*) озвучили планы к 2030-му году иметь порядка одного миллиона физических кубит, что эквивалентно порядка тысячи логических кубит. Для криптографии это еще не страшно, но вот многие полезные задачи уже можно будет попробовать решать. Ну и стоит еще раз посмотреть на график закона Мура для классических компьютеров, которые каждые десять лет показывают примерно такой же прогресс!

О квантовом превосходстве

Очень часто можно услышать разговоры о том, что достигнуто или опровергнуто квантовое превосходство. Попробуем под конец лекции разобраться, что же это такое и почему это важно (или не важно).

Понятие квантового превосходства

Само понятие было сформулировано еще в 2012-м году известным физиком теоретиком Джоном Прескиллом.



Fig. 9 Джон Прескил, который и придумал этот термин. Еще он известен своим знаменитым пари с другим физиком Стивеном Хокингом (которое Хокинг проиграл)

Квантовое превосходство это решение на квантовом компьютере задачи, которую нельзя решить на классическом компьютере за разумное время (10 тысяч лет разумным временем не считается). Достижение квантового превосходства это однозначно новый уровень в развитии квантовых вычислений. Но есть один подвох. Дело в том, что речь идет о совершенно любой задаче, независимо от того, насколько она полезна или бесполезна.

Так что когда кто-то заявляет о достижении квантового превосходства, то это важный повод для ученых и разработчиков квантовых компьютеров, но скорее всего это очень малозначимый факт, с точки зрения простого обывателя.

Хронология событий

Ну и в конце приводим краткую хронологию событий.

- 2019 год, компания *Google* заявляет о достижении квантового превосходства. Задача выбрана максимально удобная для квантового компьютера и полностью лишенная практического смысла. По словам разработчиков из *Google* их квантовый компьютер за 4 минуты решил задачу, которую классический суперкомпьютер решал бы 10 тысяч лет. Их квантовый компьютер имел 54 кубита;
- 2019 год, компания *IBM* заявляет, что *Google* не учли, что их задачу можно решать на классическом компьютере более оптимально, но без экспериментов;

- 2020 год, компания *Alibaba* реализует алгоритм *IBM* на своем суперкомпьютере и решает задачу за $\backslash(\backslashsim 20)$ дней;
- 2021 год, группа китайских ученых оптимизирует классический алгоритм и решает задачу на 60 видеокартах *NVIDIA* за 7 дней;
- 2021 год, группа других китайских ученых заявляет, что достигла нового превосходства на квантовом компьютере из 56 кубит;

В общем сейчас идет довольно интересный процесс войны меча и щита. Пока одни ученые строят более мощные квантовые компьютеры, другие придумывают более продвинутые алгоритмы их симуляции. Хотя конечно все ученые говорят, что уже где-то на 60-70 кубитах эта история окончательно закончится в пользу квантовых компьютеров.

А как это вообще выглядит? И сколько стоит?

На сегодня почти все известные технологии создания квантовых компьютеров требуют чего-то из:

- сверхнизкие температуры
- сверхвысокий вакум
- сверхточная юстировка лазеров на оптическом столе

Или даже всего сразу. Поэтому сегодня почти все квантовые компьютеры продаются через облачные сервисы. Например, относительно недавно ведущий поставщик облачных технологий – компания *Amazon* добавила в свой сервис [AWS](#) новый продукт [Amazon Braket](#). Этот продукт позволяет взять в аренду самый настоящий компьютер точно также, как мы привыкли брать в аренду процессоры, видеокарты или жесткие диски. Аналогичные продукты сейчас предоставляют и другие крупные игроки на рынке облачных услуг. Хотя это все пока исключительно для целей исследования. Ведь как мы уже поняли, сегодня квантовые компьютеры еще не способны решать реальные задачи. Стоит такое развлечение не очень дорого, например, можно запустить свою квантовую программу на 32-х кубитном компьютере [Aspen - 9](#) всего за \$0.3 (per-task).

Некоторые производители идут дальше и предлагают относительно компактные решения. Так, недавно [было представлено 24-х кубитное решение](#), которое помещается в две стандартных серверных стойки. Но масштабируемость таких устройств вызывает вопросы.

В любом случае, в ближайшие 15-20 лет точно не стоит ждать появление карманного квантового компьютера, или хотя бы квантового сопроцессора в домашнем ПК. Да и в этом нет особого смысла, ведь мало кому дома нужно взламывать биткоин, решать логистическую проблему или разрабатывать высокотемпературный сверхпроводник.

Заключение

Это вводная лекция, она не даст вам каких-то особых знаний. Скорее, ее цель заинтересовать читателя. Самое интересное будет в основной части курса, где мы будем разбирать квантовые алгоритмы, пытаться симулировать квантовую механику и обучать самые настоящие квантовые нейросети! Ждем вас на курсе!

Пререквизиты для прохождения курса QML

Настоящий курс совмещает в себе как теорию для глубинного понимания материала, так и большое количество практики для закрепления навыков и проработки неочевидных ситуаций/проблем. Поэтому предполагается, что некоторые базовые знания у вас уже есть. Проверить себя можно с помощью optionalного [входного теста](#), который поможет понять, нужно ли проходить вводные "голубые" лекции – по Python, математике и машинному обучению (15 вопросов всего, по 5 на блок). Тест довольно простой, так что если вы его прошли без ошибок или почти без ошибок – это мало о чём говорит, все равно дальше может быть сложно. А вот если ошибок много, тогда вам стоит обратиться ко вводным "голубым" блокам.

Python, Poetry и среда разработки

Для успешного прохождения курса вы должны уметь делать и понимать такие вещи, как:

- копирование файлов с GitHub;
- установка [Poetry](#) и библиотек [Python](#);
- как запустить [Jupyter Notebook](#);
- основные сочетания клавиш в [Jupyter Notebook](#);
- как запускать код в [Jupyter Notebook](#);
- какие простые типы данных существуют в [Python](#);
- как работать со сложными типами-контейнерами ([list](#), [dict](#), [tuple](#)), включая самые популярные методы каждого из них;
- как работает арифметика [Python](#) и как переменные разных типов соотносятся друг с другом;
- что такое условные конструкции и ветвление логики;
- что такое циклы и как их можно использовать применительно к сложным типам данных;
- как писать функции в [Python](#), что такое аргументы функции и какими они бывают;
- что такое декоратор и как его использовать;
- базовый ООП: объявление классов, понятие объекта класса, умение писать методы класса.

Если какая либо из тем вызывает у вас непонимание, то настоятельно рекомендуется освежить знания (или даже их приобрести), пройдя соответствующие уроки в блоке "Основы Python". Даже если у вас есть опыт работы с ним и написания кода, то предлагается ознакомиться с процессом установки всех необходимых зависимостей и библиотек (урок 2 блока), а также горячими клавишами и разнообразными функциями в [Jupyter Notebook](#) (урок 3).

Линейная алгебра и [NumPy](#)

Для понимания базовых квантовых алгоритмов и алгоритмов квантового машинного обучения, как и в классическом машинном обучении, не обойтись без линейной алгебры. Если есть время взяться за математику по-серьезному, рекомендуем курсы [Gilbert Strang](#) или [Д.В. Беклемишева](#), а также [задачник А.И. Кострикина](#). А если Вы это когда-то изучали линейную алгебру, но подзабыли ее, то поможет блок нашего курса. Хочется, чтобы у Вас не было проблем с такими понятиями как:

- Вектор-строка и вектор-столбец
- Линейная зависимость векторов
- Системы линейных уравнений
- Матрицы, определитель и след матрицы
- Линейный оператор
- Комплексное и эрмитово сопряжение
- Скалярное и векторное произведения
- Гильбертово пространство

Список явно не исчерпывающий и скорее всего явно так и не перечислишь все то, что может встретиться по мере изучения квантовых алгоритмов.

Машинное обучение

Наконец, чтобы говорить о квантовом машинном обучении, надо иметь неплохое представление о собственно машинном обучении. [Блок нашего курса](#) рассказывает о том, что общего у разных задач машинного обучения и как из основных компонентов, подобно пазлу, складываются различные применения моделей машинного

обучения. Этот блок, впрочем, не заменяет хорошей книги или курса по машинному обучению (как [курс Andrew Ng](#), [специализация Яндекса и МФТИ](#) или [mlcourse.ai](#)), покрывающего такие темы как:

- обучение с учителем и без учителя, обучение с подкреплением
- метрики качества в задачах обучения с учителем
- функции потерь и сведение задач машинного обучения к задачам оптимизации
- обработка и построение признаков
- оценка качества модели и кросс-валидация
- алгоритмы классификации и регрессии: деревья решений, лес, бустинг, логистическая регрессия и др.
- алгоритмы ранжирования, метрики качества ранжирования
- обучение без учителя: снижение размерности, кластеризация, поиск аномалий
- байесовский взгляд на машинное обучение
- обучение представлению

Опять же, это не исчерпывающий список, к тому же покрывающий только “классику”, игнорируя глубокое обучение. Но если что-то из перечисленного вызывает недоумение или если со входным [текстом](#) не задалось, возможно, стоит обратиться к упомянутым курсам.

О блоке “Python”

Этот блок включает в себя:

- описание процесса установки среды разработки для прохождения курса;
- описание синтаксиса языка программирования Python;
- разбор типовых синтаксических конструкций при программировании на Python;
- примеры анализа возникающих ошибок, помогающие в дальнейшем прохождении курса.

Вводная лекция про Python

Добро пожаловать во вводный блок курса! Если у вас есть опыт программирования на [Python](#), рекомендуем ознакомиться с пререквизитами курса – так вы сможете понять, какие разделы стоит изучить или освежить в памяти перед прохождением курса. Тест можно игнорировать, если вы только начинаете свое знакомство с [Python](#).

Итак, [Python](#) – это такой язык программирования, который позволяет сообщить компьютеру о том, что нужно сделать, чтобы достичь некоего результата. За последнее десятилетие он получил быстрое распространение и сейчас является одним из самых популярных языков программирования в мире. Входной порог для его использования достаточно низок: вы можете использовать [Python](#) для решения своих задач даже если никогда не имели дела с программированием.

Что такое [Python](#)?

[Python](#) – это язык программирования общего назначения, используемый во многих приложениях. Например:

- разработка веб-приложений;
- создание игр;
- продвинутая аналитика данных, в том числе с использованием нейронных сетей;
- компьютерная графика;
- геофизика;
- психология;
- химия;
- теория графов.

[Python](#) используют практически все крупные компании, о которых вы слышите каждый день: Google, Yandex, YouTube, Dropbox, Amazon, Facebook ... список можно продолжать часами.



Fig. 10 Рассчитанная в Python симуляция преломлений света черной дырой

Чем примечателен Python?

В основе разностороннего применения и популярности лежит **простота изучения**: все чаще люди начинают свой путь в программировании с **Python**, поскольку он очень **дружелюбен к новичкам** и позволяет максимально быстро перейти к решению целевой задачи.

Сюда же можно отнести **многообразие библиотек** (или *расширений функциональности*, то есть кода, написанного другими людьми, который вы можете переиспользовать). Хотите изучить физику небесных тел и симулировать их взаимодействия? Можно найти и скачать библиотеку, позволяющую за один вечер провести вычисления, о которых в прошлом веке можно было лишь мечтать. Хотите создать прототип мобильного приложения? И на этот случай есть библиотека. Вам нравится квантовая физика и вы хотите использовать ее вместе с умными компьютерными алгоритмами? Что ж, тогда вы снова по адресу.

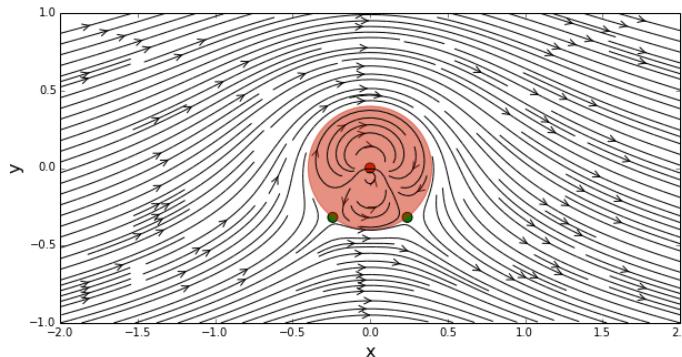


Fig. 11 Пример моделирования аэродинамики в Python с помощью библиотеки AeroPython

Python – это **высокоуровневый язык для быстрой разработки и/или прототипирования**, на нем очень удобно проверять гипотезы и идеи. “**Высокоуровневый**” означает, что вам не нужно вникать в устройство компьютера и тонкости взаимодействия с ним, чтобы перейти к задаче. Многое «сделано за вас»: вы работаете с простыми **абстракциями** (или удобными представлениями), а не боретесь с компьютером из-за непонимания сложностей его устройства.

Еще один плюс в копилку популярности языка – это **элегантность и краткость синтаксиса** (принципов написания кода, как будто это абзацы в тексте или колонки в газете). Вместе с вышеупомянутым обилием библиотек вы можете буквально за 5 минут и 10 строк кода – а это меньше половины листа А4 – воспроизвести научную статью, в которую вложено несколько человеко-лет. А еще такой синтаксис делает **код легким для чтения, запоминания и понимания**.

Столицей отметить, что **Python** – это **интерпретируемый** язык, а значит, компьютер каждый раз перед выполнением программы читает код строчку за строчкой и определяет (интерпретирует), что нужно сделать дальше, не проводя никаких оптимизаций и предварительных расчетов. Это негативно влияет на общую скорость работы: **Python** является одним из самых медленных языков. Тем не менее он отлично подходит для академических целей, например, исследовательской работы или других задач, где скорость работы не является критически важной. Настоящая сила **Python** заключается в том, что это “**язык-клей**”: он обеспечивает удобный доступ к различным библиотекам, написанным на высокоеэффективных языках, например, на C/C++, Fortran, CUDA C и других.

И в чем подвох?

В простоте языка и его доступности для быстрого старта таится одна из проблем: вы можете не понимать, что происходит внутри, поэтому иногда бывает сложно разобраться в причинах ошибок и неточностей, возникающих по ходу работы над задачей. В целом к Python применим следующий принцип: "Easy to learn, hard to master". Возвращаясь к примеру элегантности кода, когда 10 строк кода выполняют всю работу: важно понимать, что за ними стоят еще сотни или даже тысячи строк кода, а это может приводить к ситуациям, когда поиск ошибки в минимальном наборе команд растягивается на несколько дней.

Но не пугайтесь!

В данном блоке мы постараемся дать вам всю необходимую интуицию и теорию для успешного прохождения настоящего курса, ответим на основные вопросы, покажем типовые примеры использования Python и разберем классические ошибки.

Дополнительно отметим, что Python хорош и для квантового машинного обучения (QML), ради которого весь курс и затянут, в особенности – для классического машинного обучения (ML). В области ML этот язык программирования стал де-факто стандартом, который используют практически все специалисты.

Интересные факты

Еще немного дополнительной информации про Python:

- Разработчик языка Гвидо ван Россум назвал его в честь популярного британского комедийного телешоу 1970-х «Летающий цирк Монти Пайтона».
- Актуальной версией Python считаются версии 3.6 и выше (3.7, 3.8.12...). Долгое время (до 2020 года) существовал Python 2, который ныне не поддерживается и не обновляется. Если вы видите кусок кода на Python 2 и вам предстоит работать с ним, возможно, сначала придется его переписать, хотя большая часть кода имеет совместимость и работает корректно. В этом курсе мы не будем изучать Python 2.
- У Python огромное сообщество: большинство проблем, с которыми вы можете столкнуться, уже было озвучено и даже решено. Это означает, что используя поисковик, вы можете решить практически все проблемы в течение 10-30 минут. Главное – научиться правильно формулировать свои вопросы.
- При работе с Python следует придерживаться принципа "должен существовать один и, желательно, только один очевидный способ сделать это". Другие принципы (Дзен Питона) на русском языке – [по ссылке](#).
- Python – это открытый проект, в который каждый может внести изменения (но они должны быть предварительно одобрены), например, [тут](#).
- Есть целый набор рекомендаций и предложений по улучшению кода ([PEP](#), [Python Enhancement Proposals](#)). Они содержат указания на то, как следует писать код и чего стоит избегать, а также дискуссии о будущих изменениях в языке.
- Язык постоянно развивается, в нем появляются новые возможности, улучшается производительность (скорость выполнения).
- Сборник всех существующих в открытом доступе библиотек [находится тут](#).
- Если вы столкнетесь с багом (системной ошибкой, вызванной внутренним механизмом языка), то сообщить об этом можно [на специальном сайте](#).



Fig. 12 Кот для привлечения внимания и в благодарность за то, что вы начали проходить курс и сделали самый сложный шаг – прошли первую лекцию!

Установка необходимых инструментов и настройка среды

Описание лекции

Эта лекция расскажет:

- как скачать курс себе;
- как собрать свой экземпляр книги;
- как скачать сотню библиотек за раз;
- как проверить, корректно ли прошла установка.

Git и получение файлов курса

[Git](#) - это система контроля версий. Этот инструмент часто используется для совместной разработки программ (и не только!) группой людей, каждый из которых работает над своей отдельной проблемой. Так, например, этот курс был создан с помощью [Git](#), и каждый автор работал с отдельной копией, создавая набор лекций. А затем все копии собирались в одну книгу, которую вы и читаете. Для вашего удобства, а также возможности получить (“собрать” на сленге) свой экземпляр книги (если он вам зачем-то нужен – ведь все доступно онлайн в полном объеме). К сожалению, знакомство с этим инструментов не входит в программу настоящего курса, и единственное, что вам потребуется - это скачать файлы себе для дальнейшей работы. Хороший туториал на английском можно найти по [этой ссылке](#).

Однако сейчас ограничимся лишь простой операцией: для этого надо зайти на страницу с репозиторием (папкой [по ссылке на github](#) (одну из разновидностей [Git](#)). Затем найдите зеленую кнопку с текстом `<> Code` и кликните по ней. В открывшемся окне убедитесь, что выбрана вкладка [HTTPS](#), а не [SSH](#) или [GitHub CLI](#). В самом низу окошка будет кнопка [Download ZIP](#) - она нам и нужна, жмите!

Tip

Если вы проследуете инструкциям из документации по установке [Git](#), то действия выше можно проделать одной командой из консоли: `git clone https://github.com/SemyonSinchenko/qmlcourse`.

Вообще [Git](#) - очень полезный инструмент, и рекомендуется уделить время прочтению инструкций по его установке и настройке, это пригодится в будущем.



Fig. 13 Не пугайтесь темного цвета - это просто цветовая схема в браузере. На светлой странице все, что написано выше, применимо в полном объеме.

Затем скачанный файл необходимо разархивировать любым удобным способом. В современных версиях Windows можно выбрать соответствующий пункт в выпадающем меню при клике на скачанный архив правой кнопкой мыши, на [MacOS](#) достаточно просто кликнуть дважды, а на [Ubuntu](#) ... что вы делаете на этой странице? Нужно поставить программу, если ее еще нет, а затем разархивировать по любому удобному пути: `sudo apt install unzip && unzip qmlcourse-master.zip -d <путь/к/вашей/папке/>`.

Теперь в папке `qmlcourse-master` вы увидите содержимое курса, над которым трудилась наша команда: это и картинки, и текст, и задания.

Что делать дальше

В этом блоке мы сделаем две вещи: соберем книгу и установим на ваш компьютер все необходимое для прохождения курса, от [Python](#) и его библиотек до специальных программ, которые занимаются отслеживанием этих самых библиотек. Буквально после пары строк на вашем компьютере появится сотня библиотек, которыми мы хвастались в первой лекции – они позволят как проводить математические расчеты, так и писать программы для квантовых компьютеров. Итак, дальнейшие шаги зависят от вашей операционной системы:

Инструкция для Windows

Для дальнейшей работы необходимо установить менеджер пакетов для Windows - [chocolatey](#). Официальная инструкция находится [по этой ссылке](#). В данной инструкции рассмотрена установка без администраторских прав, то есть вы можете установить все программы даже на рабочий ноутбук или компьютер - дополнительных соглашений или паролей для установки не требуется.

Итак, необходимо перейти по ссылке выше и скопировать код файла [ChocolateyInstallNonAdmin.ps1](#). Создайте где-нибудь (хоть на рабочем столе) пустой файл с указанным именем, вставьте в него скопированное содержимое.

Tip

Важно, чтобы расширение файла было именно `ps1`. Если вы не знаете, как создавать файлы произвольных расширений, то картинка ниже поможет.



Fig. 14 Необходимо зайти в "Проводник" (любая папка с документами), затем перейти во вкладку "Вид", после чего поставить галочку у соответствующего пункта. Теперь при создании файла, к примеру, `my_file.abc` расширение будет `abc`, а не `txt` (хотя файл может оставаться текстовым). Такой файл можно открыть через блокнот (нажав правой кнопкой мыши и выбрав программу для запуска).

Если вы все сделали правильно, то иконка файла изменится:



Fig. 15 Иконка изменилась (она соответствует [PowerShell](#), см. ниже). Файл можно редактировать через блокнот, как показано выше.

Поздравляем, вы создали исполняемый файл! В своем роде это скрипт, который и запустит установку. Для этого нужно открыть [PowerShell](#) (терминал в рамках операционной системы Windows). Его вы можете найти очень просто - достаточно набрать название в меню "Пуск". После этого выполните команду (скопируйте, вставьте и нажмите [Enter](#)):

```
Set-ExecutionPolicy Bypass -Scope Process -Force;
```

После этого перетащите в терминал файл (прямо мышкой с рабочего стола) и отпустите левую кнопку: в терминале должен появиться путь до файла. Снова выполните команду ([Enter](#)). Дождитесь окончания установки.

И, наконец, закончим установкой вспомогательных инструментов. Запустите команду

```
choco install make curl miniconda3
```

Все сопроводительное обеспечение установлено, осталось непосредственно собрать окружение и собственный экземпляр книги - в этом нам поможет запуск двух команд (**но перед этим нужно перейти в директорию со скачанным курсом**. Для этого нужно воспользоваться командой `cd` (change directory, сменить директорию) и указать путь к директории. Самый простой способ это сделать - перетащить папку из окна проводника с файлами (или рабочего стола) в терминал: путь подставится автоматически. Пример правильной команды: `cd /home/ivan.petrov/Downloads/course/qmlcourse-master/`):

```
make install-windows  
make build-windows
```

Инструкция для MacOS

Для дальнейшей работы необходимо установить менеджер пакетов для MacOS - `brew`. [По ссылке](#) вы найдете инструкцию и более детальное описание, чем является `brew`. Вам необходимо выполнить в терминале (который доступен как приложение на экране со всеми приложениями – просто начните вводить `term` и поиск подскажет верный вариант) команду:

```
/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"
```

⚠ Attention

Скорее всего, для выполнения этой команды, нужно задействовать команду `sudo`, это вполне нормальная история.

Затем выполнить простую команду (тоже в терминале), предварительно перейдя в папку с курсом, которую вы извлекли из архива. Для этого нужно воспользоваться командой `cd` (change directory, сменить директорию) и указать путь до папки. Самый простой способ это сделать - перетащить папку из окна Finder в терминал: путь подставится автоматически. Например:

```
cd /home/ivan.petrov/Downloads/course/qmlcourse-master/
```

И из этой папки вызовите последовательность из двух команд:

```
make install-macos-latest  
make build-linux-macos
```

Инструкция для Ubuntu

Эта часть подойдет также для всех “родственных” систем: Debian, Linux Mint, Deepin Linux (который можно встретить на ноутбуках компании Huawei, к примеру) и многие другие системы, использующие менеджер пакетов `apt`. Для начала необходимо установить дополнительные программы. Выполните код ниже в терминале:

```
sudo apt-get update  
sudo apt-get install build-essential -y  
sudo apt-get install git wget -y
```

⚠ Attention

При использовании `sudo` терминал может попросить вас ввести пароль для установки и обновления пакетов программ. Это абсолютно正常но.

Затем выполнить простую команду (тоже в терминале), предварительно перейдя в папку с курсом, которую вы извлекли из архива. Для этого нужно воспользоваться командой `cd` (change directory, сменить директорию) и указать путь до папки. Самый простой способ это сделать - перетащить папку из окна Finder в терминал: путь подставится автоматически. Например:

```
cd /home/ivan.petrov/Downloads/course/qmlcourse-master/
```

И из этой папки вызовите последовательность из двух команд:

```
make install-ubuntu-latest  
make build-linux-macos
```

💡 Hint

Первая команда установит **Python**, **poetry** - это еще один менеджер, который отвечает за управление библиотеками **Python** и их версиями, а также сами библиотеки, необходимые для прохождения курса; вторая – соберет **jupyter book**, то что по ссылке https://semyonsinchenco.github.io/qmlcourse/_build/html/book/index.html (так что можете не выполнять если не хотите, трава и так будет расти)

Проверка корректности установки

Чтобы убедиться в том, что все отработало штатно, и у вас полный комплект, запустим очень мощный инструмент для работы в машинном обучении - **Jupyter Notebooks**. С ним вы познакомитесь ближе в следующей лекции, а пока введите в терминале команду

```
poetry run jupyter notebook
```

💡 Hint

Для того, чтобы запустить что-либо в текущем окружении, вам нужно начинать писать команду с **poetry run** (как в примере выше), ибо все ваши установленные библиотеки (которые в свою очередь устанавливаются как **poetry add <name_lib>**) не будут подгружаться

В терминале должна появиться надпись о том, что запущен сервер, а также будет предоставлено две или три ссылки.



```
> jupyter notebook
[I 01:33:33.945 NotebookApp] The port 8888 is already in use, trying another port
.
[I 01:33:33.945 NotebookApp] The port 8889 is already in use, trying another port
.
[I 01:33:34.098 NotebookApp] JupyterLab extension loaded from /Users/igor.kotenko/v/anaconda3/lib/python3.8/site-packages/jupyterlab
[I 01:33:34.098 NotebookApp] JupyterLab application directory is /Users/igor.kotenkov/v/anaconda3/share/jupyter/lab
[I 01:33:34.100 NotebookApp] Serving notebooks from local directory: /Users/igor.kotenkov
[I 01:33:34.100 NotebookApp] Jupyter Notebook 6.1.4 is running at:
[I 01:33:34.100 NotebookApp] http://localhost:8890/?token=3ff7308b48d2fc2da2c077f39e209b1aa01b62ce4c65dae
[I 01:33:34.101 NotebookApp] or http://127.0.0.1:8890/?token=3ff7308b48d2fc2da2c077f39e209b1aa01b62ce4c65dae
[I 01:33:34.101 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[I 01:33:34.101 NotebookApp]

To access the notebook, open this file in a browser:
file:///Users/igor.kotenkov/Library/Jupyter/runtime/nserver-41145-open.html
Or copy and paste one of these URLs:
http://localhost:8890/?token=3ff7308b48d2fc2da2c077f39e209b1aa01b62ce4c65dae
or http://127.0.0.1:8890/?token=3ff7308b48d2fc2da2c077f39e209b1aa01b62ce4c65dae
```

Fig. 16 Примерно так будет выглядеть терминал со ссылками.

Перейдите по одной из ссылок, в которых есть слово **token** – они начинаются с <http://localhost> или <http://127.0.0.1>. В вашем браузере откроется страница с обзором папки, из которой был запущен **Jupyter Notebook**. Если вы в дальнейшем хотите хранить все свои результаты в другой папке, то перед запуском команды с помощью уже указанной инструкции в терминале **cd <путь/до/директории>** перейдите к ней.

Для создания нового файла – блокнота, как его еще называют (почему? узнаете в следующей лекции) – кликните по кнопке **new** в правом верхнем углу, а затем – по **Python 3** (возможно будет какая-нибудь приставка, это штатно).



Fig. 17 Таким образом можно создать новый файл с кодом.

Для завершения проверки скопируйте код ниже в тетрадку в браузере, а затем нажмите **CTRL + Enter** или **command + Enter** (это заставит код выполниться, подробнее дальше в курсе). Если увидите график - то все в полном порядке!

```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

# Fixing random state for reproducibility
np.random.seed(19680801)

# Compute pie slices
N = 20
θ = np.linspace(0.0, 2 * np.pi, N, endpoint=False)
radii = 10 * np.random.rand(N)
width = np.pi / 4 * np.random.rand(N)
colors = plt.cm.viridis(radii / 10.)

ax = plt.subplot(111, projection='polar')
ax.bar(θ, radii, width=width, bottom=0.0, color=colors, alpha=0.5)

plt.show()
```

Не переживайте, этот код не нужно разбирать сейчас – мы просто убеждаемся, что все работает согласно задумке. Если что-то не так, пересмотрите все ли сделали согласно инструкции; если да и не воспроизводится, то задавайте вопрос в [канале курса](#)

Что мы узнали в лекции

- что такое **Git** и как скачать проект на локальный компьютер;
- как становить необходимые библиотеки, **Python** и прочие прелести;
- как запустить **Jupyter Notebook** для набора кода;

Знакомство с инструментарием: Jupyter

Описание лекции

Эта лекция расскажет:

- что такое **Jupyter** и чем он хорош;
- о видах ячеек и режимах работы в **Jupyter**;
- о самых необходимых горячих клавишах;
- что такое ядро и почему вас это должно волновать.

Введение

Как любой мастер должен знать свой инструмент, так и любой человек, решивший пройти курс QML, должен понимать тонкости рабочей среды. Как вы уже могли понять по прошлому занятию, всю (или большую часть) работы мы будем делать в **Jupyter Notebook**. Но бояться нечего: по сути, **Jupyter** – это *продвинутый текстовый редактор* с функцией запуска кода и получения результатов вычислений. Настолько продвинутый, что позволяет вам не только рисовать картинки и писать формулы, но даже строить целые интерактивные карты:



Fig. 18 Пример интерактивной визуализации прямо внутри рабочего файла с кодом

На курсе, конечно, работы с гео-данными не будет, однако очень пригодятся вывод информации в виде **таблиц**, **создание и отрисовка простых графиков**, а главное, все это происходит в **браузере**. Редактирование кода в браузере не вызывает лишних проблем со средами разработки и в то же время оно доступно максимально широкому кругу людей. В этом редакторе можно запускать **Python**-код, что очень похоже на интерактивный редактор MATLAB, если вы с ним знакомы.

Благодаря удобству использования и доступности **Jupyter** в настоящее время стал крупным игроком в нише научных вычислений и быстрого прототипирования. Вдобавок он безумно удобен для обучения и передачи знаний. Почему? Давайте разбираться.

Типы ячеек

В **Jupyter** существует несколько типов **ячеек**, мы поговорим о двух основных: **Code** и **Markdown**. В прошлом уроке мы создали пустой **ноутбук**, чтобы проверить установку **Jupyter**. **Ноутбуком** это называется потому, что в переводе с английского **notebook** – это тетрадка (альтернативное название на русском языке). В тетрадке можно писать что-то осмысленное, черкаться, оставлять пометки. Сейчас вы должны видеть вот такой экран:



Fig. 19 Пример пустой только что созданной тетрадки.

Здесь верхней красной чертой выделено поле с **названием** ноутбука. Можете кликнуть по нему, переименовать во что-то осмысленное и нажать **Enter**, чтобы применить изменения. Нижней же чертой обозначен выпадающий список переключений **типов ячеек**. По умолчанию создана одна **code** ячейка – в ней в будущем мы будем писать **Python**-код. Попробуйте кликнуть по списку и выбрать **Markdown** – визуально ячейка немного изменится.

Что такое **Markdown**?

Markdown (произносится маркдаун) – облегченный язык разметки, созданный с целью обозначения **форматирования** в простом тексте с максимальным сохранением его читаемости человеком.

Пример: `Text attributes italic, bold, `monospace`.`

С помощью **Markdown** можно разнообразить код, вставить формулы (в том числе в **LaTeX** формате, если он вам знаком), ссылки на статьи и многое другое. Попробуйте скопировать код из примера выше в **Markdown**-ячейку и нажать кнопку **Run**:



Fig. 20 Пример выполненной ячейки. Красным квадратом выделена кнопка Run.

Произошло следующее: ваша ячейка выполнилась и **Jupyter** отобразил ее содержимое. С помощью такого форматирования можно писать целые статьи с выкладками, формулами, графиками, то есть сопроводительной информацией. Поэтому, как уже было сказано, тетрадки очень удобны, особенно если соблюдать структуру, то есть писать сверху вниз с разделением на логические блоки. Также стоит отметить, что создалась новая Code-ячейка прямо под первой.

Multivariate Normal Distribution Equation

Recall the equation for the normal distribution from the **Gaussians** chapter:

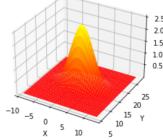
$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}(x - \mu)^2/\sigma^2\right]$$

Here is the multivariate normal distribution in n dimensions.

$$f(\mathbf{x}, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right]$$

The multivariate version merely replaces the scalars of the univariate equations with matrices. If you are reasonably well-versed in linear algebra this equation should look quite manageable. If not, don't worry, both FilterPy and SciPy provide functions to compute it for you. Let's ignore the computation for a moment and plot it to see what it looks like.

```
In [14]: import kf_book.mkf_internal as mkf_internal
mean = [2., 17.]
cov = [[10., 0.],
       [0., 4.]]
mkf_internal.plot_3d_covariance(mean, cov)
```



This is a plot of multivariate Gaussian with a mean of $\mu = [2, 17]$ and a covariance of $\Sigma = [10, 0; 0, 4]$. The three dimensional shape shows the probability density for any value of (X, Y) in the x -axis. I have projected the variance for x and y onto the walls of the chart - you can see that they take on the Gaussian bell curve shape. The curve for X is wider than the curve for Y , which is explained by $\sigma_x^2 = 10$ and $\sigma_y^2 = 4$. The highest point of the 3D surface is at the means for X and Y .

Fig. 21 Пример грамотно оформленной Jupyter-тетрадки. Такую можно скинуть коллегам – и всем все будет понятно!

Note

Самая прекрасная часть тетрадок: **ячейки разных типов можно смешивать по порядку**, таким образом сначала описывая какую-то логику, а затем непосредственно реализовывая ее в коде и выполняя.

Вам не так часто придется писать Markdown-заметки самостоятельно, основная причина их создания – ваше желание и дальнейшее удобство использования ноутбука. Есть также другая причина, по которой мы акцентируем на них внимание. Может так произойти, что вы случайно изменили тип ячейки и не заметили этого. Теперь, если в Markdown-ячейку вставить **Python**-код, то ничего не произойдет или возникнет ошибка. Если вы заметили что-то странное при выполнении кода в тетрадке – **проверьте, корректен ли тип ячейки**. Для выполнения кода нужно выставить тип **Code**.

```
In [1]: print('Привет!')
Привет!
```

```
In [ ]: print('Привет!')
```

Fig. 22 Верхняя ячейка с `print` написана в Code-режиме и корректно выполняется, печатая строку

с приветствием. Нижняя ячейка же содержит текст, а не код, поэтому работать не будет

(точнее, код отобразится как текст, но не будет выполнен).

**Tip**

Обратите внимание, как визуально отличаются эти две ячейки. Одна из них имеет прозрачный фон, другая – серый. У **Code** ячейки также есть странная надпись слева (про нее еще поговорим).

Каждый раз вручную запускать код (или Markdown) через кнопку **Run** не очень-то удобно, поэтому можно запомнить две комбинации клавиш. **CTRL+Enter** выполнит текущую ячейку и оставит “курсор” (указатель на ячейку) на том же месте, не создавая лишнюю строчку в ноутбуке. **Shift+Enter** повторит функциональность кнопки **Run**: выполнит ячейку, а затем перейдет на следующую (или создаст новую, если текущая ячейка является последней).

Первая комбинация (**CTRL+Enter**) будет полезна в том случае, если вы что-то написали и знаете, что будете вносить изменения (например, менять цвет линии на графике в попытках добиться визуальной красоты), а значит, придется менять код в этой же ячейке.

Вторая (**Shift+Enter**) пригодится тогда, когда вы хотите запустить много-много идущих подряд ячеек (можете представить, что коллега скинул вам свою тетрадку с 30 клетками и вы хотите ее запустить, чтобы получить данные).

Не беспокойтесь, буквально к концу первого блока лекций у вас выработается мышечная память и вы будете использовать сочетания клавиш на автомате.

Режимы работы

Пришло время разобраться с цветом курсора, выделяющего ячейки. Он может быть **синим** или **зеленым**.

```
In [5]: print('Еще одна ячейка!')
Еще одна ячейка!
```

```
In [5]: print('Привет!')
Привет!
```

```
In [5]: print('Еще одна ячейка!')
Еще одна ячейка!
```

Fig. 23 Пример разного цвета указателя клетки.

Никакой тайны за этим нет, это два режима: **режим редактирования** и **командный режим**. Зеленый цвет сигнализирует о том, что вы работаете с **текстовым содержимым ячейки**, то есть редактируете его! Можете писать код, вставлять формулы, что угодно. Но как только вы нажмете **ESC** на клавиатуре, цвет сменится на синий, что означает **возможность редактирования всего ноутбука, а не отдельных ячеек в нем**. Можно передвигать ячейки, удалять их (полностью, а не только текст в них), добавлять новые. Стрелочками на клавиатуре можно выбирать ячейки (скакать вверх и вниз). Как только доберетесь до нужной (а вместо этого можно просто кликнуть по ней мышкой, что полезно в ситуации, когда клетка очень далеко, в самом низу страницы) – жмите **Enter**, чтобы вернуться к редактированию.



Tip

Можно осуществлять переходы между режимами кликом мышки (внутри блока кода либо где-нибудь в стороне, слева или справа от ячейки, где ничего нет).



Fig. 24 Кнопки управления ноутбуком.

Выполнять описанные выше операции можно с помощью горячих клавиш (или хоткеев), либо через интерфейс. Описание выделенного блока кнопок для картинки выше (в порядке слева направо, с указанием сочетаний клавиш):

1. Сохранение ноутбука (**CTRL+S**) – делайте его **пощаще**, чтобы не потерять результаты работы!
2. Создание ячейки ниже текущей (**B**) – **B** потому, что создается клетка снизу, то есть **Below**. Логика для **A** и **Above** аналогична.
3. Вырезать ячейку (**X**) – применимо и к целому блоку ячеек (можно выделить с зажатой клавишей **Shift**). Функциональность как и в Excel/Word: убрать в одном месте, чтобы вставить в другом.
4. Копировать ячейку (**C**).
5. Вставить ячейку из буфера (**V**) – после вырезания или копирования ячейки.
6. Переместить текущую выделенную ячейку **вверх**.
7. Переместить текущую выделенную ячейку **вниз**.

Описание всех доступных команд (и соответствующих им хоткеев) доступно при нажатии на кнопку с клавиатурой в правой части выделенного блока меню (вне красного прямоугольника).



Попробуйте потратить 5-7 минут на практику использования этих кнопок и сочетаний клавиш.

Первое время можете пользоваться только элементами UI-интерфейса – это нормально, главное, сопоставить кнопки и стоящую за ними функциональность.

Оставшиеся кнопки на панели

Про кнопку **Run** (и хоткей **Shift+Enter**) мы уже поговорили, а что с остальными?



Для того, чтобы вы могли запускать код **Python**, запускается так называемое **“ядро” (или kernel)**, то есть приложение, которое непосредственно выполняет (запускает) ваш код и передает результаты обратно в **Jupyter**-ноутбук. За это отвечает как раз **Run**.

Справа от нее расположен **Stop**, который **прерывает выполнение программы**. Он может быть полезен в случаях, когда вы запустили расчеты на час, но заметили ошибку – и поэтому нужно и код переписать, и ячейку с кодом снова запустить. И в этой ситуации вы сначала **останавливаете** выполнение, редактируете код, затем жмете **Run** – и все готово!

Но случается беда и код не останавливается, потому что ядро **Python** зависает. В таких случаях нужно **перезапустить ядро** – и кнопка с закругленной стрелочкой **Restart** поможет осуществить задуманное. Будьте аккуратны – **вы потеряете ВСЕ несохраненные данные** (значения переменных, результаты расчетов, данные для построения графиков). Сама тетрадка останется без изменений, то есть **написанное сохранится**. Концепция “ядра” и запуска кода станет более понятна, когда мы перейдем к практике.

Tip

Пока стоит держаться правила: “Накосячил? Попробуй остановить (**Stop**) ядро. Не получается? Тогда перезапускай (**Restart**) его!”

Нужно понимать, что ядро “помнит” все предыдущие выполненные ячейки (пока не будет перезагружено или выключено), а значит, вы можете позже в коде переиспользовать те части, которые были описаны ранее (например, переменные или физические константы). Иными словами, состояние ядра сохраняется во времени и между ячейками – **оно относится к документу в целом**, а не к отдельным ячейкам.

Последняя кнопка из выделенного блока имеет говорящее название: **Re-start and run all**. Ядро будет **перезапущено** (все переменные и данные удалятся), а затем **каждая ячейка будет выполнена в порядке сверху вниз**. Поэтому рекомендуется соблюдать структуру, чтобы запускать код с нуля (после возвращения к ноутбуку на следующий день, но с новым ядром, так как компьютер был выключен) – и он отрабатывал.

Что это за In [*]?

Та самая надпись слева от запущенной Code-ячейки. Это вспомогательная информация о том, что происходит с **кодовой** ячейкой (**In** означает **Input**, то есть ввод кода). Возможно несколько вариантов заполнения.

```

In [ ]: print('Я еще не запущена()')
In [6]: print('Я уже не запущена()')
Я уже не запущена()

In [*]: while True:
         continue
         print('А я в бесконечном цикле...')

```

Fig. 26 Пример трех видов информации о статусе ячейки.

В первом случае в квадратных скобках **ничего нет** – это значит, что **ячейка еще не была запущена**. Возможно, вы забыли, а быть может, она просто ждет своего часа.

Во втором случае **ячейка была запущена** шестой по счету (да-да, **ячейки выполняются по порядку, который задаете вы сами!**) и она успешно отработала и завершилась.

В последней строчке умышленно был сделан бесконечный цикл. Это означает, что код *никогда не сможет выполниться* и будет висеть до тех пор, пока вы не остановите (**Stop**нете) ядро. Поэтому там выведен **индикатор выполнения ячейки** – в скобках указана звездочка *****. Обратите внимание: **это не всегда плохой сигнал**. Если ваш код должен выполняться 2-3 минуты, то все это время будет выводиться **[*]**. Когда код отработает и результат будет получен, отрисуется цифра (например, **[7]**).

Самая полезная клавиша

Пришло время программировать! Скопируйте себе в ноутбук кусок кода ниже и попробуйте его запустить. Не переживайте, он может показаться сложным и непонятным, но сейчас не требуется понимание всех деталей.

```

import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

# Fixing random state for reproducibility
np.random.seed(19680801)

# Compute pie slices
N = 20
theta = np.linspace(0.0, 2 * np.pi, N, endpoint=False)
radii = 10 * np.random.rand(N)
width = np.pi / 4 * np.random.rand(N)
colors = plt.cm.viridis(radii / 10.)

ax = plt.subplot(111, projection='polar')
ax.bar(theta, radii, width=width, bottom=0.0, color=colors, alpha=0.5)
plt.show();

```

Вы увидите ошибку. По сообщению видно (стрелочка в левой части указывает на проблемное место), что во второй строке используется слово **mat**, при этом **Python** жалуется на отсутствие такого модуля. Все дело в том, что в коде выше производится попытка рисования графика и для этого используется библиотека **matplotlib**. Но в одной из строк написано только **mat**. Это не дело, давайте исправлять. Однако всех библиотек не запомнишь – и это не нужно. Попробуйте поставить курсор после буквы **t** (и перед точкой) и нажать **TAB**. Вы должны увидеть **список подсказок** и из него выбрать нужный вариант. Этот список не только сокращает время написания кода (за счет автоматического дополнения), но и позволяет избежать ошибок в написании. Обязательно пользуйтесь этим инструментом.

Если вы все сделали правильно, воспользовавшись подсказкой, то после очередного запуска (**Run**) кода появится рисунок.

```
In [17]: the_answer_to_the_ultimate_question_of_life_the_universe_everything = 42  
the_answer_but_divided_by_two = 21  
  
In [ ]: the_an|  
the_answer_but_divided_by_two  
In [ ]: the_answer_to_the_ultimate_question_of_life_the_universe_everything
```

Fig. 27 Другой пример удачного использования: есть несколько переменных со сложным, но очень похожим названием. Не стоит их перепечатывать – достаточно нажать TAB!

Что ж, большое количество новой информации позади, давайте подведем итоги!

Что мы узнали из лекции

- Существует два основных типа ячеек, один предназначен для программирования (написания кода), другой – для описания (формул, определений).
- Существует два режима работы: **edit** – редактирование конкретной ячейки и **control** – работа со всей структурой тетрадки/ноутбука.
- Режимы работы можно различить по цвету рамки вокруг активной ячейки, а тип ячейки – по прозрачности фона и наличию надписи In[] слева.
- Между режимами можно переключаться с помощью Esc и Enter.
- Чтобы запустить ячейку (с кодом или текстом), нужно нажать на кнопку Run сверху, либо воспользоваться сочетаниями клавиш: Shift+Enter, CTRL+Enter.
- Нужно не забывать сохранять ноутбук (CTRL+S), а быстро добавить ячейку кода можно с помощью плюсика слева сверху (или клавиши B).
- Клавиши C / V позволяют копировать и вставлять ячейки в control-режиме.
- In[3] указывает на порядок выполнения ячеек с кодом, In[*] – на процесс выполнения.
- Если вы долго ждете выполнения Code-блока, можно Stopнуть ядро, если не помогает – Restartнуть.
- Ядро – это процесс, выполняющий код, и после перезагрузки оно не сохраняет переменные.
- TAB – ваш друг, позволяющий избегать опечаток, а также реже пользоваться документацией.

Бонус-материал

Полная официальная документация по Jupyter находится по [ссылке тут](#).

Если вам захочется узнать больше о трюках в ноутбуках, о недостатках и преимуществах по сравнению с альтернативами, предлагаем посмотреть выступление Joel Grus:



Переменные и вывод информации в Python

Описание лекции

Из этой лекции вы узнаете о:

- базовых типах данных и переменных;
- простейших операциях с числами и строками;
- функции print;

- выводе информации с подстановкой значений.

Суть переменных в Python

Настало время приступить к изучению непосредственно **Python**, ведь прошло три лекции, а мы об языке программирования и не говорили вовсе! И поскольку наш курс посвящен физике, то начнем со [знакомой всем по школьным карандашам формулы](#) ($E=mc^2$). По ней можно вычислить полную энергию физического объекта (E) с помощью известной массы объекта (m) и константы (c). Эта постоянная, указывающая на скорость света в вакууме, используется настолько часто, что для нее выделили **отдельное обозначение в виде буквы латинского алфавита**, как и для многих других аналогичных величин. Если в формуле встречается (c) (в известном контексте), то вы всегда уверены, что именно нужно подставить при расчетах.

Этот пример полностью описывает концепцию **переменных** в языках программирования, и **Python** не исключение. Запись ($x = 3$) означает, что везде по тексту далее под иксом подразумевается именно тройка, и ничего другого (пока не будет введено новое определение). Этой же логике подчиняется **Python**. Сначала указывается **имя переменной**, а затем – ассоциируемое с ней значение.

```
c = 299_792_458 # запишем константу, м/c
m = 0.5 # масса некоторого абстрактного объекта, кг
E = m * (c ** 2) # вычисляем энергию, дж

some_variable_1 = 10.2 # какая-то другая переменная
m = 12
```

Пример кода выше иллюстрирует сразу несколько базовых концепций, которые нужно запомнить:

1. В объявлении переменной нет ничего сложного. Синтаксис и правила интуитивно понятны: это можно делать как в физике/математике, как в учебниках и статьях.
2. **#** означает комментарий, то есть произвольный текст, который не воспринимается **Python** (все **до конца строки** кода полностью игнорируется). Служит исключительно для создания подсказок в коде, объяснения происходящего, то есть для удобства.
3. Числа могут быть **целыми и вещественными**. Разряды в целых числах для удобства визуального восприятия можно разделять нижней чертой.
4. **Значение переменной может быть вычислимым**, то есть являться производной от других переменных (как (E), ведь это результат перемножения). На самом деле значение вычисляется в момент объявления переменной (при сложной формуле расчета процесс может занимать некоторое время).
5. Операция возведения в степень реализуется с помощью **`**`**.
6. В качестве названия переменных можно использовать **буквы и цифры**, а также некоторые символы. Однако **имя переменной не может начинаться с цифры**.
7. Переменные можно переопределять (и даже менять тип). Однако **старое значение в этом случае будет безвозвратно утрачено**. В данном примере после выполнения последней строчки нельзя установить, чему было равно (m) до того, как переменной было присвоено значение дюжины.

Если говорить менее строго и более абстрактно, то **переменная – это контейнер** (или коробка), в котором что-то лежит, и на самой коробке на приклеенном листочке бумаги указано содержимое. Чем понятнее надпись, тем легче найти и использовать объект (поэтому переменные с названием из одной буквы воспринимаются плохо, особенно если таких переменных очень много).



Fig. 28 [Объявить переменную – значит положить объект в коробку с подписью](#).

Типы переменных

В листинге кода выше важно заметить, что существует разница между двумя типами численных переменных: **целые и вещественные**. При сугубо математических расчетах и арифметических операциях тип переменной не имеет значения. Однако для некоторого функционала в **Python** нужно быть аккуратным. Мы поговорим подробно об этом в следующих лекциях, а пока стоит запомнить, что вещи, которые необходимо посчитать – в том числе и **длину** чего-то счетного – должны быть целочисленными (как и в жизни: первый, второй, третий...).

⚠ Attention

Целочисленный тип называется **int** (от **Integer**), вещественный – **float**. Эти типы можно переводить из одного в другой. При переводе вещественного числа в целое теряется часть информации.

Тип переменной – и это относится не только к числам, но и к **любому** объекту – можно узнать с помощью функции **type**. Для вывода информации в **Python** используется функция **print**. Что именно представляет собой функция мы рассмотрим в более поздних лекциях, пока стоит думать об этом как о некотором объекте, который зависит (рассчитывается) от других объектов и выдает некоторый результат. Для передачи аргументов используются круглые скобки (аналогично математике: $(y = F(x))$). Давайте скомбинируем эти знания и рассмотрим пример:

```
first_variable = 10
second_variable = 10.0

# запишем в переменные значения типов данных
type_of_first_variable = type(first_variable)
type_of_second_variable = type(second_variable)

# и распечатаем сами типы, чтобы посмотреть глазами и сравнить
print(type_of_first_variable)
print(type_of_second_variable)

# перезапишем переменные
first_variable = 12.9
second_variable = int(first_variable)
third_variable = float(second_variable)

# в print() можно передавать несколько переменных
print(first_variable, second_variable, third_variable)
```

```
<class 'int'>
<class 'float'>
12.9 12 12.0
```

Внимательно проанализируйте код выше – в нем продемонстрирован базовый синтаксис **преобразования типов и вывода информации**. Легко увидеть подтверждение высказанных ранее тезисов: **second_variable** действительно потеряла часть информации (дробную часть числа), которую нельзя вернуть, если преобразовать переменную обратно во **float**. Преобразование типов в языках программирования называется **приведением** (типов, то есть привести одно к другому, а не из-за страшилок про духов).

Арифметические операции с числами

Математика **Python** максимально близка к естественной: **+-*** и ****** (рассмотренное ранее возведение в степень) работают в точности как ожидается. С делением **/** есть нюанс: **возвращаемое значение всегда вещественное**.

```
a = 3
b = 12.1

c = a + b

# можно объединять вызовы функций print и type
# без создания лишней переменной
print(type(c))

# и даже трех функций, включая приведение типа
print(type(int(c)))

# деление числа на само себя дает единицу, но..
print(a / a)
print(b / b)
print(c / c)
print(12 / 4)
```

```
<class 'float'>
<class 'int'>
1.0
1.0
1.0
3.0
```

Note

Обратите внимание, что операции не изменяют переменную саму по себе (то есть операция `a + b` не меняет ни `a`, ни `b`). Чтобы сохранить получаемое значение, нужно присвоить его некоторой переменной (в примере выше это `c`). Если вы хотите изменить непосредственно саму переменную, то можно переприсвоить ей значение на основе расчета: `a = a + b` или `c = c + 12`.

Даже несмотря на то, что кейс с делением числа на само себя очевиден (всегда получается единица, кроме деления на нуль), будет выведено вещественное значение. Сами же вещественные значения можно складывать, вычитать, умножать и возводить в степень как с целыми, так и с вещественными числами (и наоборот). Если в таком выражении используется хотя бы одна `float`-переменная, то и результат будет не целочисленным. Однако:

```
a = 3
b = 2

print(a + b, type(a + b))
print(a * b, type(a * b))
print(a ** b, type(a ** b))
```

```
5 <class 'int'>
6 <class 'int'>
9 <class 'int'>
```

Это *практически* все тонкости, которые необходимо знать, чтобы не совершать базовые ошибки.

Примечание

Возможно, у вас родился вопрос относительно расстановки пробелов в коде выше. Обязательно ли соблюдать такой синтаксис? Нужно ли ставить пробелы до и после знаков операций? На самом деле нет: это делается исключительно для удобства чтения кода и **настоятельно рекомендуется не удаляться от стандартов языка**. Код ниже выполнится без ошибок, однако ухудшается читаемость:

```
a=      3
b      =2

print(a +b, type(a+ b))
print(a      * b, type(a *b))
print(a**b, type(a      ** b))
```

```
5 <class 'int'>
6 <class 'int'>
9 <class 'int'>
```

Строковые переменные

Мы разобрались в том, как описывать и хранить числа, как производить арифметические расчеты. Базовый математический язык освоен, но мы же люди, и хочется общаться словами! Конечно, `Python` позволяет это делать. Благодаря **строковым переменным** можно хранить и соединять текстовую информацию:

```
text_variable = 'тут что-то написано'
another_text_variable = "Вася, впиши сюда что-нибудь перед публикацией курса!"

long_text = """
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor
incididunt ut
labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation
ullamco laboris
nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in
voluptate velit
esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non
proident, sunt
in culpa qui officia deserunt mollit anim id est laborum.
"""

print(another_text_variable)
```

Вася, впиши сюда что-нибудь перед публикацией курса!

В примере выше рассмотрено три способа создания текстовых переменных. Первые два не отличаются между собой с точки зрения [Python](#), то есть неважно, используете ли вы одинарные кавычки ' или двойные ". Однако стоит понимать, что если ваша строка содержит в себе такой символ, то кавычка должна быть изменена:

```
error_string = 'Chillin' kid'
another_error_string = "И тут он мне говорит: "у тебя нет ног!""

print(error_string)
print(another_error_string)
```

Механизм ошибки таков, что [Python](#) неясно: это вы закончили строчку и дальше идет какая-то команда, или же строчка продолжается. В обоих случаях **нужно сменить способ создания строки** – и тогда все будет хорошо:

```
error_string = "Chillin' kid"
another_error_string = 'И тут он мне говорит: "у тебя нет ног!"'

print(error_string)
print(another_error_string)
```

Chillin' kid
И тут он мне говорит: "у тебя нет ног!"

Если необходимо сохранить какой-либо объемный текст или сообщение, можно воспользоваться мультистрочным объявлением переменной, как в первом примере блока.

Строки можно объединять для удобства вывода информации:

```
first_string = 'Результат вычислений: '
second_string = ". Это не так много!"

a = 12
b = 2
result = a * b

# два способа вывода:
print(first_string, result, second_string)

# либо через склейку строк вручную
# обратите внимание на приведение типа int к str
result_string = first_string + str(result) + second_string
print(result_string)
```

Результат вычислений: 24 . Это не так много!
Результат вычислений: 24. Это не так много!

⚠ Warning

Будьте аккуратны со сложением строк. Объединение строк "3" и "5" даст результат "35", а не 8 – и тип результирующего значения **будет строковый**. **Сложить строку и число нельзя**: вы получите ошибку и никакого приведения типов не произойдет. Здесь возникнет двусмысленность – нужно привести число к строке и затем сконкатенировать или же строку к числу (а вдруг это невозможно?), после чего сложить.

Склейивание строк называется **конкатенацией**. Попробуйте в **Jupyter**-ноутбуке объединить строковые, целочисленные и вещественные переменные в разных комбинациях. Разберитесь, что означает ошибка, которая будет выведена в случае, если не делать приведение типов (то есть без `str` в `str(result)`).

Note

Обратите внимание на пробел между числом и точкой в первом случае. Они добавлены автоматически функцией `print` – это сделано для того, чтобы разные объекты при последовательном выводе не “склеивались” друг с другом. Во втором случае этого не происходит, так как мы напрямую склеиваем строки и только затем передаем результат конкатенации на печать в `print`.

Но на практике это не совсем удобно, поэтому в **Python** придумали **F-строки**. Их суть в том, что переменная из кода напрямую подставляется (с автоматическим приведением типа к строке) в саму строку! Вот:

```
a = 12
b = 2
result = a * b

result_string = f'Результат вычислений: {result}. Это не так много!'

# и без f
wrong_result_string = 'Результат вычислений: {result}. Это не так много!'

print(result_string)
print(wrong_result_string)
```

```
Результат вычислений: 24. Это не так много!
Результат вычислений: {result}. Это не так много!
```

Для объявления **F-строки** нужно, во-первых, использовать одинаковые кавычки на концах текста. Во-вторых, нужно указать литеру `f` перед самой строкой. И последнее – нужно обрамить название конкретной переменной (`result` в данном случае) в фигурные скобки.

Когда переменная одна, а также нет текста после ее использования, то выгода **F-строк** не так очевидна (относительно простого `print(some_string, some_variable)`). Однако представьте, что вам нужно вывести координаты точки в трехмерном пространстве, значение времени, параметры системы и значение некоторой функции от всех переменных выше!

```
# так тоже можно!
x, y, z = 12.1, 0, 13
# скобки, как и в математике, задают порядок выполнения вычислений
func_val = (x * y) ** z

current_time = 30.113412

# а вот так можно писать длинные f-строки (но работает и для обычных)
out_string = (f'В точке с координатами X={x}, Y={y}, Z={z} значение функции '
             f'равно {func_val}. Состояние системы указано на момент '
             'времени t=' + str(current_time))

print(out_string)
```

```
В точке с координатами X=12.1, Y=0, Z=13 значение функции равно 0.0. Состояние
системы указано на момент времени t=30.113412
```

Что мы узнали из лекции

- **Переменные – это “контейнеры”,** в которые можно что-то положить и дать название.
- Математика в **Python** не имеет сложных правил, процесс вычислений максимально интуитивен.
- Арифметические операции могут **менять тип** результирующей переменной.
- `type()`, `print()` – базовые функции, с помощью которых можно делать **самопроверки** по ходу написания кода.
- **Сменить тип** переменной можно вызовом функций `int()`, `float()`, `str()`.
- Строки могут обрамляться как `'`, так и `"` (но этих символов **не должно быть внутри** текста).
- **F-строки** облегчают комплексный вывод, содержащий как текст, так и переменные **Python** (и автоматически приводят типы).

Условные конструкции, булева логика и сравнения

Описание лекции

В этой лекции мы расскажем про:

- `if/else`-конструкции и условия;
- тип `bool`;
- операторы сравнения;
- блоки кода и отступы.

Ветвление логики

В прошлых лекциях мы рассмотрели программы с линейной структурой: сначала выполнялась первая конструкция (например, объявление переменной), затем вторая (преобразование переменной или расчет по формуле), после – третья (`print` для вывода результатов). Можно сказать, что происходило последовательное исполнение команд, причем каждая инструкция выполнялась **обязательно**. Но что делать, если хочется опираться на обстоятельства и принимать решения о том, выполнять одну часть кода или другую?

Допустим, по числу `x` нужно определить его абсолютную величину, то есть модуль. Программа должна напечатать значение переменной `x`, если $x > 0$ или же величину $-x$ в противном случае ($-(-5) = 5$). Эту логику можно записать следующим образом:

```
x = -3 # попробуйте поменять значение переменной

if x > 0:
    print("Исходный x больше нуля")
    print(x)
else:
    print("Исходный x меньше или равен нулю")
    print(-x)
```

```
Исходный x меньше или равен нулю
3
```

В этой программе используется условная инструкция `if` (в переводе с английского “если”). `if` – это ключевое зарезервированное слово (так нельзя называть свою переменную), указывающее на **условную конструкцию**. После `if` следует указать вычислимое выражение, которое **можно проверить на истинность** (то есть можно сказать, правда это или нет). Общий вид конструкции следующий:

```
if (Условие):
    <Блок инструкций 1>
else:
    <Блок инструкций 2>
```

`else` – тоже ключевое слово (в переводе – “иначе”). Таким образом, можно в голове придерживаться такой интерпретации: “**если** условие верно (истинно), **то** выполнни первый блок команд, **иначе** выполнни второй блок”.

Условная инструкция содержит как минимум ключевое слово `if` (единожды), затем может идти любое количество (включая ноль) блоков с условием `else if <условие>` (иначе если, то есть будет выполнена проверка нового условия в случае, если первая проверка в `if` не прошла), затем – дополнительно – конструкция `else`. Логика чтения и выполнения кода сохраняет порядок **сверху вниз**. Как только одно из условий будет выполнено, выполнится соответствующая инструкция (или набор инструкций), а все последующие блоки будут проигнорированы. Это проиллюстрировано в коде:

```
x = -3.8 # попробуйте поменять значение переменной

if x > 0:
    print('x больше нуля')
elif x < 0: # можно написать "else if x < 0:"
    print('x меньше нуля')
else:
    print('x в точности равен нулю')
print('Такие дела!')
```

```
x меньше нуля  
Такие дела!
```

Понятно, что `x` не может одновременно быть и больше нуля, и меньше (или равен ему). Среди всех трех `print`-блоков будет выполнен **только один**. Если `x` действительно больше нуля, то второе условие (`x < 0`) даже не будет проверяться – **Python** сразу же перейдет к последней строке и выведет надпись “Такие дела!”.

Чтобы лучше разобраться в том, как работает код, можно использовать **визуализаторы** – например, [такой](#). Прогоняйте через него весь код (даже в несколько строк) и сверяйте со своими ожиданиями от его работы.

А что вообще такое эти ваши условия?

Выше было указано, что после конструкций `if/else if` необходимо указать **условие**, которое еще и должно быть истинным или ложным (“правда или нет”). Давайте попробуем определить необходимый **тип** переменной.

```
x = -3.8  
  
condition_1 = x > 0  
condition_2 = x < 0  
  
print(condition_1, type(condition_1))  
print(condition_2, type(condition_2))
```

```
False <class 'bool'>  
True <class 'bool'>
```

Видно, что оба условия имеют один и тот же тип – `bool`, то есть **boolean**. По [определению](#):

Boolean (Булев, Логический тип данных) – примитивный тип данных в информатике, которые могут принимать **два возможных значения**, иногда называемых истиной (True) и ложью (False).

Оказывается, что в коде выше мы получили **ВСЕ** возможные варианты булевой переменной – это истина (`True`, пишется только с заглавной буквы) и ложь (`False`, аналогично). Никаких других значений быть для условия не может. Вот такой это простой тип данных.

Способы получения `bool`

Какими вообще могут быть условия? Как с ними можно обращаться? Согласно [официальной документации](#), в **Python** есть такие операторы сравнения:

Операция	Значение
<code><</code>	строго меньше чем
<code><=</code>	меньше или равно
<code>></code>	строго больше, чем
<code>>=</code>	больше или равно
<code>==</code>	равный
<code>!=</code>	не равный

Fig. 29 Все операции сравнения работают нативно (так же, как и в математике)

```
print(3.0 > 3)  
print(3.0 == 3)
```

```
False  
True
```

Здесь практически нечего рассматривать, операторы сравнения они и в **Python** операторы. Куда интереснее принцип **объединения различных условий в одно** – для создания комплексной логики.

Пусть стоит задача определения четверти точки по ее координатам на двумерной плоскости. Решение такой задачи может быть записано следующим образом:

```
x = -3.6
y = 2.5432

if x > 0:
    if y > 0:
        # x > 0, y > 0
        print("Первая четверть")
    else:
        # x > 0, y < 0
        print("Четвертая четверть")
else:
    if y > 0:
        # x < 0, y > 0
        print("Вторая четверть")
    else:
        # x < 0, y < 0
        print("Третья четверть")
```

Вторая четверть

Пример показывает, что выполняемым блоком кода может быть любой блок **Python**, включая новый логический блок с **if-else** конструкцией. Однако его можно сократить с помощью **логических операторов and, or и not**. Это стандартные логические операторы [Булевой алгебры](#).

Логическое **И** является бинарным оператором (то есть оператором с двумя operandами: левым и правым) и имеет вид **and**. Оператор **and** возвращает **True** тогда и только тогда, когда **оба его операнда имеют значение True**.

Логическое **ИЛИ** является бинарным оператором и возвращает **True** тогда и только тогда, когда **хотя бы один операнд равен True**. Оператор "логическое ИЛИ" имеет вид **or**.

Логическое **НЕ** (отрицание) является унарным (то есть **с одним операндом**) оператором и имеет вид **not**, за которым следует единственный операнд. Логическое **НЕ** возвращает **True, если операнд равен False и наоборот**.

Эти правила необходимо запомнить для успешного создания сложных условий с целью разделения логики, заложенной в **Python**-коде.

Проиллюстрируем правила в коде на простых примерах. Обратите внимание на то, как можно объявлять **bool**-переменные – это не сложнее, чем создание целочисленного значения:

```
true_value = True
false_value = False

# False потому, что один из operandов является False
some_value = true_value and false_value
print(some_value)

# True потому, что хотя бы один из operandов равен True
some_value = true_value or false_value
print(some_value)

# отрицание True (истины) есть False (ложь)
some_value = not true_value
print(some_value == false_value)

# пример сложного условия - порядок лучше в явном виде задавать скобками
hard_condition = (not true_value or false_value) or (true_value != false_value)
print(hard_condition)
```

False
True
True
True

Теперь попробуем их применить на приближенных к практике примерах:

```

x = -3.6
y = 2.5432

if x > 0 and y > 0: # конструкция заменяет два вложенных if
    print("Первая четверть")
elif x > 0 and y < 0:
    print("Четвертая четверть")
elif y > 0:
    print("Вторая четверть")
else:
    print("Третья четверть")

# определим, большое ли число x (в терминах модуля)
x_is_small = (x < 3) and (x > -3)
# число большое, если оно не маленькое (по модулю)
x_is_large = not x_is_small # можно отрицать факт малости x

print('Is x small? ', x_is_small)
print('Is x large? ', x_is_large)

# так тоже можно писать - на манер неравенств в математике
another_x_is_small = -3 < x < 3
print(another_x_is_small)
print(another_x_is_small == x_is_small)

```

```

Вторая четверть
Is x small? False
Is x large? True
False
True

```

Так как вторая переменная `x_is_large` – это отрицание (`not`) первой (`x_is_small`), то они **никогда** не будут равны.

Блоки кода и отступы

В примерах выше вы наверняка заметили упоминание термина “блок кода”, а также откуда-то взявшимся отступы после условий, и это не случайно. Во-первых, давайте признаем, что так условные конструкции (особенно вложенные!) читать куда легче, и глаза не разбегаются. Во-вторых, это особенность языка **Python** – здесь не используются скобки `{}` для указания блоков, все форматирование происходит с помощью отступов. Отступы **всегда** добавляются в строки кода **после двоеточия**.

Для выделения блока инструкций (строк кода, выполняющихся подряд при любых условиях), относящихся к инструкциям `if`, `else` или другим, изучаемым далее, в языке **Python** используются **отступы**. Все инструкции, которые относятся к одному блоку, должны иметь **равную величину отступа**, то есть одинаковое число **пробелов в начале строки**. В качестве отступа [PEP 8](#) рекомендует использовать **отступ в четыре пробела** и не рекомендует использовать символ табуляции. Если нужно сделать еще одно вложение блока инструкций, достаточно добавить еще четыре пробела (см. пример выше с поиском четверти на плоскости).

Tip

Хоть и не рекомендуется использовать символ табуляции для создания отступов, кнопка `Tab` на вашей клавиатуре в **Jupyter**-ноутбуке (при курсоре, указывающим на начало строки кода) создаст отступ в четыре пробела. Пользуйтесь этим, чтобы не перегружать клавишу пробела лишними постукиваниями :).

Что мы узнали из лекции

- Для задания логики выполнения кода и создания нелинейности используются **условные инструкции**, поскольку они следуют некоторым условиям.
- Условная инструкция задается ключевым словом `if`, после которого может следовать несколько (от нуля) блоков `else if/elif`, и – дополнительно – в конце добавляется `else`, если ни один из блоков выше не сработал.
- Условия должны быть **булевого типа** (`bool`) и могут принимать **всего два значения** – `True` и `False`. Выполнится тот блок кода, который задан истинным (`True`) условием (и только первый!).
- Условные конструкции можно вкладывать друг в друга, а также объединять с помощью **логических операторов** `and`, `or` и `not`.
- **Блок кода** – это несколько подряд идущих команд, которые будут выполнены последовательно.
- Чтобы выделить блок кода после условия, используйте **отступы** – четыре пробела.
- Чтобы создать отступ в **Jupyter**, нужно нажать `Tab` в начале строки кода.

Списки и циклы в Python

Описание лекции

На этом занятии мы разберем следующие темы:

- списки (`list`) и их методы;
- индексация списков;
- что такое срезы и зачем они нужны;
- цикл `for` и функция `range`;
- итерация по спискам, list comprehensions.

Введение в списки объектов

В предыдущих лекциях мы оперировали малым количеством переменных. Для каждого блока логики или примера кода вводилось 3-5 объектов, над которыми осуществлялись некоторые операции. Но что делать, если объектов куда больше? Скажем, вам необходимо хранить информацию об учащихся класса – пусть это будет рост, оценка по математике или что-либо другое. Не знаю, как вы, но я нахожу крайне неудобным создание 30 отдельных переменных. А если еще и нужно посчитать среднюю оценку в классе!

```
average_grade = petrov_math + kosareva_math + zinchenko_math + kotenkov_math + ...
average_grade = average_grade / 30
```

Такой код к тому же получается крайне негибким: если количество студентов, как и их состав, изменится, то нужно и формулу переписать, так еще и делитель – в нашем случае 30 – изменять.

Часто в программах – даже в (квантовом) машинном обучении – приходится работать с большим количеством однотипных переменных. Специально для этого придуманы **массивы** (по-английски array). В Python их еще называют **списками** (`list`). В некоторых языках программирования эти понятия отличаются, но не в Python. Список может хранить переменные **разного** типа. Также списки называют "**контейнерами**", так как они хранят какой-то набор данных. Для создания простого списка необходимо указать квадратные скобки или вызвать конструктор типа (`list` – это отдельный тип, фактически такой же, как `int` или `str`), а затем перечислить **объекты через запятую**:

```
# различные способы объявления списков
first_list = []
second_list = list()
third_list = list([1,2, "stroka", 3.14])
fourth_list = [15, 2.2, ["another_list", False]]

print(type(second_list), type(fourth_list))
print(first_list, fourth_list)
```

```
<class 'list'> <class 'list'>
[] [15, 2.2, ['another_list', False]]
```

Tip

Хоть список и хранит переменные разного типа, но так делать без особой необходимости не рекомендуется – вы сами скорее запутаетесь и ошибетесь в обработке объектов списка. В большинстве других языков программирования массив может хранить только объекты одного типа.

Для хранения сложных структур (скажем, описание студента – это не только оценка по математике, но и фамилия, имя, адрес, рост и так далее) лучше использовать классы – с ними мы познакомимся в будущем. А еще могут пригодиться **кортежи**, или `tuple`. Про них в лекции не рассказано, самостоятельно можно ознакомиться [по ссылке](#).

Теперь можно один раз создать список и работать с ним как с единственным целым. Да, по прежнему для заведения оценок студентов придется разово их зафиксировать, но потом куда проще исправлять и добавлять! Рассмотрим пример нахождения средней оценки группы, в которой всего 3 учащихся, но к ним присоединили еще 2, а затем – целых 5:

```
# базовый журнал с тремя оценками
math_journal = [3, 3, 5]

# добавим новопришедших студентов
math_journal.append(4)
math_journal.append(5)

# и сразу большую группу новых студентов
math_journal.extend([2,3,4,5,5])

print(f"math_journal = {math_journal}")

# найдем среднюю оценку как сумму всех оценок, деленную на их количество
avg_grade = sum(math_journal) / len(math_journal)
print(f"avg_grade = {avg_grade}")
```

```
math_journal = [3, 3, 5, 4, 5, 2, 3, 4, 5, 5]
avg_grade = 3.9
```

В коде выше продемонстрировано сразу несколько важных аспектов:

1. Добавлять по одному объекту в конец списка можно с помощью метода списка `append`;
2. Метод `append` принимает в качестве аргумента один Python-объект;
3. Слияние списков (конкатенация, прямо как при работе со строками) нескольких осуществляется командой `extend` (расширить в переводе с английского);
4. Для списков определена функция `len`, которая возвращает целое число `int` – количество объектов в списке;
5. Функция `sum` может применяться к спискам для суммирования всех объектов (если позволяет тип – то есть для `float`, `int` и `bool`). Попробуйте разобраться самостоятельно, как функция работает с последним указанным типом);
6. Для методов `append` и `extend` не нужно приравнивать результат выполнения какой-то переменной – изменится сам объект, у которого был вызван метод (в данном случае это `math_journal`);
7. Списки в Python **упорядочены**, то есть объекты сами по себе места не меняют, и помнят, в каком порядке были добавлены в массив.

Tip

В тексте выше встречается термин **метод**, который, быть может, вам не знаком. По сути метод – это такая же **функция**, о которых мы говорили раньше, но она принадлежит какому-то объекту с определенным типом. Не переживайте, если что-то непонятно – про функции и методы мы поговорим подробно в ближайших лекциях!

`print`, `sum` – функции, они существуют сами по себе; `append`, `extend` – методы объектов класса `list`, не могут использоваться без них.

Индексация списков

Теперь, когда стало понятно, с чем предстоит иметь дело, попробуем усложнить пример. Как узнать, какая оценка у третьего студента? Все просто – нужно воспользоваться **индексацией** списка:

```
# базовый журнал с пятью оценками
math_journal = [1, 2, 3, 4, 5]

third_student_grade = math_journal[3]
print(third_student_grade)
```

4

И снова непонятный пример! Давайте разбираться:

1. Для обращения к `i`-тому объекту нужно в квадратных скобках указать его индекс;
2. **Индекс** в Python начинается **С НУЛЯ** – это самое важное и неочевидное, здесь чаще всего случаются ошибки;
3. Поэтому `[3]` обозначает взятие **четвертой** оценки (и потому выводится четверка, а не тройка);
4. Всего оценок 5, но так как индексация начинается с нуля, то строчка `math_journal[5]` выведет ошибку – нам доступны лишь индексы `[0, 1, 2, 3, 4]` для взятия (так называется процедура обращения к элементу списка по индексу – взятие по индексу).



Fig. 30 Пример списка из трех объектов. Сверху показаны их индексы, включая отрицательные

Также в **Python** существуют отрицательные индексы (-1, -2 ...). Они отсчитывают объекты списка, начиная с конца. Так как нуль уже занят (под первый объект), то он не используется.

```

# базовый журнал с пятью оценками
math_journal = [1, 2, 3, 4, 5]

# возьмем последнюю оценку
last_grade = math_journal[-1]
print(f"Последняя оценка: {last_grade}")

# а теперь -- предпоследнюю
prev = math_journal[-2]
print(f"Предпоследняя оценка: {prev}")

# конечно, взятие по индексам можно использовать в ранее разобранном синтаксисе

if math_journal[-1] < math_journal[-2]:
    math_journal[-1] += 1
    print("Последняя оценка меньше предпоследней. Натянем студенту?")
else:
    math_journal[-2] = 2
    print("Последний студент сдал очень хорошо, на его фоне предпоследний просто
двоечник!")

```

```

Последняя оценка: 5
Предпоследняя оценка: 4
Последний студент сдал очень хорошо, на его фоне предпоследний просто двоечник!

```

Все это важно не только для грамотного оперирования конкретными объектами, но и следующей темы -

Срезы

Срезы, или slices – это механизм обращения сразу к нескольким объектам списка. Для создания среза нужно в квадратных скобках указать двоеточие, слева от него – индекс начала среза (по умолчанию 0, можно не выставлять) **включительно**, справа – границу среза **не включительно** (пустота означает “до конца списка”). Может показаться нелогичной такая разнородность указания границ, но на самом деле она безумно удобна – особенно вместе с тем, что индексация начинается с нуля. Быстрее объяснить на примере:

```

# базовый журнал с пятью оценками
math_journal = [1, 2, 3, 4, 5]

# как взять первые 3 оценки?
first_3_grades = math_journal[:3]
print(f"{first_3_grades} ")

# как взять последние две оценки?
last_2_grades = math_journal[-2:]
print(f"{last_2_grades} ")

# сделаем срез на 4 оценки, начиная со второй (с индексом 1)
start_index = 1
some_slice = math_journal[start_index : start_index + 4]
print(f"{some_slice} ")

# возьмем столько объектов из начала, сколько объектов в some_slice
yet_another_slice = math_journal[:len(some_slice)]
print(f"Верно ли, что единица входит в some_slice? {1 in some_slice}")
print(f"Верно ли, что единица входит в yet_another_slice? {1 in yet_another_slice}")

```

```
first_3_grades = [1, 2, 3]
last_2_grades = [4, 5]
some_slice = [2, 3, 4, 5]
Верно ли, что единица входит в some_slice? {1 in some_slice}
Верно ли, что единица входит в yet_another_slice? {1 in yet_another_slice}
```

Tip

Можно сделать пустой срез, и тогда Python вернет пустой список без объектов. Можете проверить сами:

```
["1", "2", "3"][10:20]
```

Давайте проговорим основные моменты, которые **крайне важно понять**:

1. Так как индексация начинается с нуля (значение по умолчанию) и правая граница не включается в срез, то берутся объекты с индексами **[0, 1, 2]**, что в точности равняется трем первым объектам;
2. Срез **[-2:]** указывает на то, что нужно взять все объекты до конца, начиная с предпоследнего
3. Значения в срезе могут быть **вычислимые** (и задаваться сколь угодно сложной формулой), но должны оставаться **целочисленными**;
4. Если нужно взять **k** объектов, начиная с **i**-го индекса, то достаточно в качестве конца среза указать **k+i**;
5. Для проверки вхождения какого-либо объекта в список нужно использовать конструкцию **x_obj in some_list**, которая вернет **True**, если массив содержит **x_obj**, и **False** в ином случае;
6. Самый простой способ сделать копию списка - это сделать срез по всему объекту: **my_list[:]**. Однако будьте внимательны – в одних случаях копирование происходит полностью (по значению), а в некоторых сохраняются ссылки (то есть изменив один объект в скопированном списке вы измените объект в исходном). Связано это с типом объектов (**mutable/immutable**), подробнее об этом будет рассказано в следующей лекции. В общем, если вы работаете с простыми типами (**int/str**), то срез вернет копию, и ее изменение не затронет исходный список. Однако для хранения новых данных нужна память, поэтому при копировании десятков миллионов объектов можно получить ошибку, связанную с нехваткой памяти.

Циклы

До сих пор в примерах мы хоть и обращались к некоторым объектам, добавляли и меняли их, все еще не было рассмотрено взаимодействие сразу с несколькими. Давайте попробуем посчитать, сколько студентов получили оценку от 4 и выше. Для этого интуитивно кажется, что нужно **пройтись по всем оценкам от первой до последней**, сравнить каждую с четверкой. Для прохода по списку, или **итерации**, используются **циклы**. Общий синтаксис таков:

```
example_list = list(...)
for item in example_list:
    <> блок кода внутри цикла (аналогично блоку в if)
    ... что-то сделать с item
    <>
```

Здесь **example_list** – это некоторый итерируемый объект. Помимо списка в Python существуют и другие итерируемые объекты, но пока будем говорить о массивах.

Этот цикл работает так: указанной **переменной item присваивается первое значение из списка**, и выполняется **блок кода** внутри цикла (этот блок, напомним, определяется отступом. Он выполняется весь от начала отступа и до конца, как и было объяснено в пятой лекции). Этот блок еще иногда называют **телом цикла**. Потом переменной **item** присваивается следующее значение (второе), и так далее. Переменную, кстати, можно называть как угодно, не обязательно **item**.

Итерацией называется каждый **отдельный проход** по телу цикла. Цикл всегда повторяет команды из тела цикла несколько раз. Два примера кода ниже аналогичны:

```
math_journal = [3, 4, 5]
counter = 0

for cur_grade in math_journal:
    if cur_grade >= 4:
        counter += 1

print(f"Всего хорошистов и отличников по математике {counter} человека")
```

```

math_journal = [3, 4, 5]
counter = 0

cur_grade = math_journal[0]
if cur_grade >= 4:
    counter += 1

# не забываем менять индекс с 0 на 1, так как каждый раз берется следующий элемент
cur_grade = math_journal[1]
if cur_grade >= 4:
    counter += 1

# и с единицы на двойку
cur_grade = math_journal[2]
if cur_grade >= 4:
    counter += 1

print(f"Всего хорошистов и отличников по математике {counter} человека")

```

Всего хорошистов и отличников по математике 2 человека

Понятно, что первый кусок кода обобщается на любой случай – хоть оценок десять, хоть тысяча. Второе решение не масштабируется, появляется **много одинакового кода, в котором легко ошибиться** (не поменять индекс, к примеру).

Движемся дальше. Так как каждый элемент списка закреплен за конкретным индексом, то в практике часто возникают задачи, логика которых связана на индексах. Это привело к тому, что появилась альтернатива для итерации по списку. Функция `range` принимает аргументы, аналогичные срезу в списке, и возвращает итерируемый объект, в котором содержатся целые числа (индексы). Так как аргументы являются аргументами функции, а не среза, то они соединяются запятой (как `print(a, b)` нескольких объектов). Если подан всего один аргумент, то нижняя граница приравнивается к нулю. Посмотрим на практике, как сохранить номера (индексы) всех хорошо учащихся студентов:

```

math_journal = [4, 3, 4, 5, 5, 2, 3, 4]
good_student_indexes = []

for student_index in range(len(math_journal)):
    current_student_grade = math_journal[student_index]
    if current_student_grade >= 4:
        good_student_indexes.append(student_index)

print(f"Преуспевающие по математике студенты находятся на позициях:
{good_student_indexes}")

```

Преуспевающие по математике студенты находятся на позициях: [0, 2, 3, 4, 7]

В примере `student_index` принимает последовательно все значения от 0 до 7 включительно. `len(math_journal)` равняется 8, а значит, восьмерка сама не будет включена в набор индексов для перебора. На каждой итерации `current_student_grade` меняет свое значение, после чего происходит проверка. Если бы была необходимость пробежаться только по студентам, начиная с третьего, то нужно было бы указать `range(2, len(math_journal))` (двойка вместо тройки потому, что индексация с нуля, ведь мы перебираем индексы массива).

Выше описаны основные концепции обращения со списками. Их крайне важно понять и хорошо усвоить, без этого писать любой код будет безумно сложно. Скопируйте примеры к себе в .ipynb-ноутбук, поиграйтесь, поменяйте параметры цикла и проанализируйте изменения.

List comprehensions

Некоторые циклы настолько просты, что занимают 2 или 3 строчки. Как пример – привести список чисел к списку строк:

```

# грубый вариант
inp_list = [1,4,6,8]
out_list = []

for item in inp_list:
    out_list.append(str(item))

# list comprehension
out_list = [str(item) for item in inp_list]
print(out_list)

```

```
['1', '4', '6', '8']
```

Две части кода идентичны за вычетом того, что нижняя – с непонятной конструкцией в скобках – короче. Python позволяет в рамках одной строки произвести какие-либо простые преобразования (помним, что `str()` – это вызов функции, а значит если у вас есть сложная функция, которая делает квантовые вычисления, то ее также можно применить!). Фактически самый частый пример использования – это паттерн “**применение функции к каждому объекту списка**”.

Что мы узнали из лекции

- `list` – это **объект-контейнер, который хранит другие объекты разных типов**. Запись происходит упорядочено и последовательно, а каждому объекту присвоен **целочисленный номер, начиная с нуля**;
- для добавления одного объекта в `list` нужно использовать метод объекта `list.append`, а для расширения списка сразу на несколько позиций пригодится `extend`;
- проверить, сходит ли объект в список, можно с помощью конструкции `obj in some_list`;
- индексы **могут быть отрицательными**: `-1`, `-2` ... В таком случае нумерация начинается от последнего объекта;
- можно получить часть списка, сделав **срез** с помощью конструкции `list[start_index : end_index]`, при этом объект на позиции `end_index` не будет включен в возвращаемый список (т.е. **срез работает не включительно по правую границу**);
- часто со списками используют **циклы, которые позволяют итерироваться по объектам массива** и выполнять произвольную логику в рамках отдельного отступом блока кода;
- для итерации по индексам можно использовать `range()`;
- простые циклы можно свернуть в `list comprehension`, и самый частый паттерн для такого преобразования – это **применение некоторой функции к каждому объекту списка** (если `x` это функция, то синтаксис будет таков: `[x(item) for item in list]`).

Словари

Описание лекции

В этой лекции мы расскажем про:

- то, что такое `dict`;
- методы и примеры использования;
- хранение изменяемых и неизменяемых типов;
- хэширование.

Что такое `dict`

`dict` (от английского “*dictionary*”, словарь) – еще один тип данных в Python. Словари хранят пары **ключ: значение**. То есть в списках можно достать элемент, если указать его позицию в виде целого числа, а в словарях – тот самый **ключ**. Обратите внимание, `dict` – **неупорядоченный** тип данных, поэтому достать элемент по номеру не получится, но отображение содержимого будет **в порядке добавления** элементов. **Уникальность** ключей должна поддерживаться, чтобы всегда можно было быстро найти одно **единственно верное значение**.

В некоторых языках программирования можно встретить ассоциативные массивы – полную аналогию `dict`. Также вспомним базы данных, широко использующиеся во всем мире для хранения информации: в таблице можно установить первичный ключ, который уникально идентифицирует запись, как и **ключ** соответствует **значению** в словаре. Самый простой пример ключа - порядковый номер объекта!

Создание словаря

Использовать словарь стоит, когда нужно сохранять объекты с какими-то ключами и обращаться к объектам по известным ключам. Один из способов определения словаря: указание пар **ключ: значение** через **запятую** внутри фигурных скобок `{}`. Напоминает `set`, правда? `{}` позволяет создать **пустой** словарь, но не пустое множество.

Например, вы решили упростить себе жизнь и больше не запоминать дни рождения коллег. Вместо этого, лучше хранить их в одном месте:

```
# локально попробуйте поменять значение переменной dates
dates = {"Кунг Фьюри": "1968-09-09", "Наташа Романова": "1985-03-15"}
print(dates)
```

```
{'Кунг Фьюри': '1968-09-09', 'Наташа Романова': '1985-03-15'}
```

В примере `dates` имеет две пары значений. В первой паре строка `"Кунг Фьюри"` является *ключом*, а `"1968-09-09"` – его *значением*.

Получение значения по ключу

Чтобы получить значение по ключу, необходимо обратиться к переменной, содержащей словарь, и указать ключ в квадратных скобках `[]`:

```
dates["Кунг Фьюри"]
```

```
'1968-09-09'
```

Если указать неверный ключ в `[]`, Python будет ругаться: выбросит ошибку `KeyError` и перестанет выполнять код. Чуть ниже посмотрим, как можно избежать таких ситуаций.

```
# пока этого ключа нет в словаре, будет ошибка при обращении
# поэтому используем перехват ошибок
try:
    print(dates["Капитан Ямайка"])

except KeyError as e:
    print(f"Ключа действительно нет: {e}")
```

```
Ключа действительно нет: 'Капитан Ямайка'
```

Изменение и добавление значений

Синтаксис изменения значения по ключу и добавления нового ключа со значением одинаковым: в `[]` нужно указать ключ, поставить `=` и указать значение, которое теперь будет соответствовать ключу.

```
# этот ключ уже был в примере
dates["Кунг Фьюри"] = "1960-09-09"

# а такого не было
dates["Капитан Ямайка"] = "1930-10-04"

print(dates)
```

```
{'Кунг Фьюри': '1960-09-09', 'Наташа Романова': '1985-03-15', 'Капитан Ямайка':
'1930-10-04'}
```

Если ключ уже был в словаре, значение по нему изменится на новое, а *старое будет удалено*. Указание нового ключа со значением добавляет пару в словарь.

Основные методы словаря

Проверка вхождения и `get()`

Помните, ранее говорили, что обращение к несуществующему ключу приводит к ошибке? Пришло время посмотреть пару способов борьбы!

Можно проверить, есть ли интересующий ключ среди множества ключей словаря. Это делается при помощи бинарного оператора `in`. Слева должен быть указан ключ, справа – переменная со словарем:

```
# еще способ создания: пары можно передавать как аргументы dict через '='
marks = dict(матан=100, английский=92)

# False
print(f'{"матан" in marks = }')

# True
print(f'{"личал" in marks = }')
```

```
"матан" in marks = False
"личал" in marks = True
```

В коде проверку можно использовать в условной конструкции `if`, чтобы принимать решение в зависимости от наличия ключа:

```
if "матан" in marks:
    print(marks["матан"])

else:
    print("Нет оценки по матану :(")
```

```
Нет оценки по матану :(
```

Теперь о методе `get()`: при помощи него тоже можно получать значения из словаря по ключу. `KeyError` никогда не появится: если ключа нет, по умолчанию возвращается `None`:

```
empty_dict = {}

# None
print(empty_dict.get("ключ"))
```

```
None
```

Вторым аргументом метода `get()` можно указать значение, которое должно возвращаться вместо `None`, когда ключ не был найден:

```
# теперь будет возвращено значение '-1'
print(empty_dict.get("ключ", -1))
```

```
-1
```

Что такое “длина словаря”?

Функция `len()` для словаря будет возвращать количество пар **ключ: значение** (их столько же, сколько ключей), которое в нем содержится:

```
# empty_dict -- пустой словарь, поэтому длина равна 0
print(f'{len(empty_dict) = }')

# а вот словарь marks уже содержит две пары, поэтому длина 2
print(f'{len(marks) = }')
```

```
len(empty_dict) = 0
len(marks) = 2
```

Удаление из словаря

Есть несколько способов очистки в словаре: можно убирать по ключу, а можно сразу удалить все!

- при помощи инструкции `del` (от английского “`delete`”) можно удалить пару **ключ: значение** (замечу, что **удаление ключа эквивалентно удалению пары `ключ: значение`**, так как мы теряем возможность найти то самое **значение**), в общем виде:

```
# таким образом из словаря "словарь" будет удален ключ "название_ключа"
# и соответствующее ему значение
del словарь[название_ключа]
```

Предположим, коллега из самого первого примера уволился и больше нет смысла хранить его день рождения:

```
# из словаря dates удаляется ключ "Наташа Романова"
del dates["Наташа Романова"]
print(dates.get("Наташа Романова"))
```

None

- **pop()** – метод, который достает значение, хранящееся по переданному ключу, и сразу удаляет ключ из словаря:

```
# еще один способ создания словаря из последовательности пар
holidays = dict([("January", [1, 2, 3, 4]), ("February", [23]), ("March", [8]))]

# pop() возвращает значение, соответствующее ключу, значит его можно присвоить
# переменной
january_days = holidays.pop("January")

# напечатается соответствующий массив
print(january_days)
```

[1, 2, 3, 4]

Для метода **pop()** есть возможность указать значение, которое будет возвращено при обращении к несуществующему ключу. Почти как **get()**, но все таки, без указания этого значения, **pop()** выбрасывает **KeyError**

- **popitem()** имеет схожее название, но не путайте с предыдущим методом: этот на вход не принимает **ключ**, а возвращает пару **ключ: значение**, которая была добавлена последней (*такое поведение гарантировано с Python 3.7*).

```
# в результате -- последняя добавленная пара
print(holidays.popitem())
```

('March', [8])

- **clear()** позволяет удалить сразу все ключи словаря, то есть полностью его очистить:

```
# вернемся к предыдущему примеру
# словарь становится пустой
holidays.clear()

# значит, длина равна 0
print(len(holidays))
```

0

Обратите внимание на то, как работают методы **pop()**, **popitem()** и **clear()**: как только вызываются, словарь меняет свой состав (*изменения происходят in place*, то изменить по месту без копирования).

Обновление и добавление ключей

Мы уже видели, что значения в словарь можно добавлять или менять, обращаясь по ключу. Python предоставляет возможность не писать кучу присваиваний, а использовать лаконичный метод **update()**, который на вход может принимать либо

- другой словарь
- пары **ключ: значение** в какой-то последовательности (например, кортежи по два значения в списке: первое – ключ, второе – значение)

Tip

Кортеж (tuple) – еще одна стандартная структура данных. Очень похожа на list (поддерживает индексацию), но является неизменяемой: как только tuple создан, поменять его нельзя

```
# создадим два словаря: в первом уже есть два ключа
quidditch_team = {"Fred Weasley": "3rd year", "George Weasley": "3rd year"}

# во втором -- один ключ
new_members = {"Harry Potter": "1st year"}

# добавим пары из new_members
# метод update() также работает in place, поэтому после выполнения данной
# строки кода, в словаре quidditch_team станет три ключа
quidditch_team.update(new_members)

print(quidditch_team["Harry Potter"])
```

```
1st year
```

А что, если в `update()` передать пары, ключ которых уже был в словаре? Значения по дублирующимся ключам будут **перезаписаны** на новые:

```
# данный ключ (то, что записано первым в кортеже) уже есть в quidditch_team
member_update = [("Harry Potter", "2nd year")]

# значение, соответствующее "Harry Potter", будет переписано
quidditch_team.update(member_update)

print(quidditch_team["Harry Potter"])
```

```
1st year
```

Доступ к ключам и значениям

В Python можно без проблем извлекать отдельно по ключам или значениям, а также итерироваться по элементам словарей в цикле `for`. Осталось разобраться, как это работает.

Ключи

По умолчанию, в конструкциях вида

```
# после in указано название переменной, хранящей словарь
for key in dict_var:
    ...
```

переменные цикла (тут – `key`) будут принимать значения из множества **ключей** словаря. Аналогично можно использовать метод `keys()` (позволяет доставать все ключи), который явно говорит, что ваш цикл идет по ключам, например:

```
# словарь в качестве ключей хранит имена игроков
for player in quidditch_team:

    # на каждой итерации будет напечатан ключ и значение
    print(player, quidditch_team[player])
```

```
Fred Weasley 3rd year
George Weasley 3rd year
Harry Potter 1st year
```

Значения

При помощи метода `values()` можно получить все значения, хранящиеся по всем ключам словаря:

```
# можно создать переменную со всеми значениями словаря
school_years = quidditch_team.values()
```

Приведем пример с циклом:

```
# словарь в качестве значений хранит годы обучения
for year in quidditch_team.values():

    # на каждой итерации будет год обучения игрока
    print(year)
```

```
3rd year
3rd year
1st year
```

Напрямую по значению получить ключ нельзя

Все и сразу

Существует метод `items()`, который достает пары **ключ: значение** в виде последовательности кортежей. Его же часто удобно использовать в циклах, чтобы не тащить длинную запись в виде названия словаря и квадратных скобок с ключом при обращении к значению:

```
# сразу две переменные: первая последовательно будет ключами,
# вторая -- значениями
for player, year in quidditch_team.items():

    # items() избавляет от необходимости обращаться quidditch_team[player],
    # чтобы получить значение. Оно уже в year
    print(f"Player {player} is in {year}")
```

```
Player Fred Weasley is in 3rd year
Player George Weasley is in 3rd year
Player Harry Potter is in 1st year
```

Сортировка

Функция `sorted()` доступна и для словарей. По умолчанию ключи словаря поддерживают порядок, в котором были добавлены, но можно отсортировать их в нужном направлении (в зависимости от типа):

```
# вспомним про рабочие дни, ключи - целые числа по номеру дня недели
week = {7: "weekend", 6: "weekend", 1: "workday"}

# в sorted_week окажутся ключи, отсортированные в порядке возрастания
sorted_week = sorted(week)
print("Порядок возрастания: {sorted_week}")
```

```
Порядок возрастания: [1, 6, 7]
```

Не забудьте, когда в функции передается просто название переменной со словарем, работа идет только над множеством ключей.

Можно ли отсортировать словарь по значениям? Да, можно попробовать самостоятельно разобраться с аргументами функции `sorted()`.

Что можно хранить

Теперь добавим немного технических подробностей: возможно, вы уже заметили самостоятельно, что `dict` может принимать в качестве ключа не всякое значение. На самом деле только **хэшируемые** объекты (можно вызвать функцию `hash()` и получить значение) могут быть ключами словаря, на значения это ограничение не распространяется. В `dict` и `set` значение хэша от объекта используется для поиска внутри структуры.

Ключом словаря нельзя сделать объект **изменяемого** типа, например, `list`, `set` или сам `dict`, так как значение их хэша может изменяться со временем. Неизменяемый кортеж может быть ключом только если не содержит внутри изменяемые объекты.

Изменяемость и неизменяемость

В англоязычной литературе изменяемые типы называют **mutable**, а неизменяемые – **immutable**, [почитать документацию](#)

В Python все – объект. Когда пользователь присваивает значение переменной, она начинает ассоциироваться с ячейкой памяти, где лежит это значение. Переменная знает адрес, откуда можно получить значение. `id()` и `hex()` показывают адрес в памяти компьютера. `_id()` – адрес в десятичном виде, а `hex()` поможет перевести в шестнадцатеричный.

По адресу лежит так называемое **внутреннее состояние** переменной:

- **неизменяемые** типы не позволяют менять внутреннее состояние, значение переменной может поменяться только вместе с адресом
- **изменяемые** типы позволяют менять внутреннее состояние переменной при сохранении адреса (возвращаемое `id()` значение не меняется, но значение переменной каким-то образом преобразовывается). Изменение по ссылке называется изменением *in place*

Неизменяемые типы

Из стандартных неизменяемыми являются:

- `int`
- `float`
- `bool`
- `str`
- `tuple`

Давайте сразу рассмотрим пример:

```
counter = 100  
# полученное вами значение адреса может отличаться  
print(counter, hex(id(counter)))
```

```
100 0x7f6f3be69d80
```



Fig. 31 `counter` указывает на 100 (цифры в нижней правой части как раз и показывают тот самый адрес)

А теперь поменяем значение `counter`:

```
counter = 200  
print(f"counter = {counter}, {hex(id(counter))} = ")
```

```
counter = 200, hex(id(counter)) = '0x7f6f3be6aa00'
```

Кажется, что раз значение переменной `counter` поменялось, то и содержимое по предыдущему адресу изменилось? Нет, на самом деле `counter` теперь указывает в другое место:



Fig. 32 `counter` указывает на новое значение 200 с другим адресом

Из интересного: Python заранее создает объекты для чисел от -5 до 256, поэтому для переменных со значением из этого диапазона берутся заранее готовые ссылки.

```
# создадим две переменные с одинаковыми значениями в диапазоне от -5 до 256
a = 20
b = 20

# a и b указывают на одно и то же место в памяти
# попробуйте у себя поменять значение a и b на число больше 256 или меньше -5
id(a) == id(b)
```

```
True
```

Изменяемые типы

Стандартные изменяемые типы это:

- `list`
- `set`
- `dict`

У списков есть метод `append()`, позволяющий добавить в него значение:

```
# создадим список и напечатаем его адрес
ratings = [1, 2, 3]
print(f"Было: {hex(id(ratings))}")

ratings.append(4)
print(f"Стало: {hex(id(ratings))} - ничего не поменялось!")
```

```
Было: 0x7f6f34b2f580
Стало: 0x7f6f34b2f580 - ничего не поменялось!
```



Fig. 33 Начальное состояние списка содержало три элемента

После добавления еще одного, адрес `ratings` не изменился.



Fig. 34 После добавления элемента поменялось лишь внутреннее состояние

Что мы узнали из лекции

- Новый тип данных – **словарь!** Позволяет хранить соответствие **ключ: значение**;
- Несколько способов создания `dict`, примеры:

```
# при помощи "литерала" - фигурных скобок {}
flowers = {"roses": "red", "violets": "blue"}

# при помощи вызова
# dict() и последовательности с парами значений
anime = dict([
    ("Ведьмак", "Кошмар волка"),
    ("Призрак в доспехах", ["Призрак в доспехах", "Синдром одиночки",
    "Невинность"]),
    []
])

# dict() и "ключ=значение"
literature = dict(poem_flowers=flowers)

print(f"\n{flowers = }\n{anime = }\n{literature = }")
```

```
flowers = {'roses': 'red', 'violets': 'blue'}
anime = {'Ведьмак': 'Кошмар волка', 'Призрак в доспехах': ['Призрак в доспехах', 'Синдром одиночки', 'Невинность']}
literature = {'poem_flowers': {'roses': 'red', 'violets': 'blue'}}
```

- методы для изменения состояния или получения доступа к элементам:

```
# доступ к элементу, если ключа нет - ошибка
print(f'{flowers["violets"] = }')

# при помощи get()
print(f'{flowers.get("magnolias") = }')
```

```
flowers["violets"] = 'blue'
flowers.get("magnolias") = None
```

```
days = ["Пн", "Вт", "Ср", "Чт", "Пт"]

# создадим пустой словарь
numbered_days = {}

# будем добавлять в него элементы в цикле
for num in range(len(days)):
    numbered_days[num] = days[num]

# получим отдельно ключи и значения
# пары из tuple можно сразу получить при помощи метода items()
keys = numbered_days.keys()
values = numbered_days.values()

print(f"Ключи: {keys}\nЗначения: {values}")
```

```
Ключи: dict_keys([0, 1, 2, 3, 4])
Значения: dict_values(['Пн', 'Вт', 'Ср', 'Чт', 'Пт'])
```

- Требование к ключу: возможность хэширования, свойство ключа внутри словаря: уникальность;
- Разобрали, что есть изменяемые (`dict, set, list`) и неизменяемые (`int, float, bool, tuple, str`) стандартные типы данных.

ФУНКЦИИ

Описание лекции

В этой лекции мы расскажем про:

- функции и их смысл;
- локальность переменных;
- параметры и аргументы;
- декораторы.

Что такое “функция” в Python

В общем случае функцией можно назвать набор связанных инструкций, которые выполняют определенную задачу. Функции во всех языках программирования помогают:

- структурировать код и улучшить читаемость
- переиспользовать код и не изобретать велосипед
- уменьшать количество мест, в которых можно ошибиться при копировании и вставке кода

В Python функции можно разделить на три типа:

- встроенные ([список built-in функций](#))
- именованные (определенные пользователем при помощи `def`)
- анонимные (`lambda`-функции)

Все функции являются объектами типа `function`.

На самом деле, вы уже использовали встроенные функции, например:

- `print()` для вывода данных на экран
- `str()` для создания объектов строкового типа
- `type()` для определения типа объекта

Ими можно пользоваться как черным ящиком, который принимает что-то на вход и творит свою магию. О том, что готовые функции ожидают получить, написано в документации вместе с описанием принципа работы.

Tip

После разбора лекции советую открыть [документацию `print\(\)`](#), например, и разобраться с подробностями работы.

Очевидно, стандартные функции дают лишь базовый инструментарий. Реализовать более сложную логику можно самостоятельно.

Пример:

Представим, что необходимо находить количество корней квадратного уравнения - это легко можно реализовать на Python!

```
def count_roots(a, b, c):
    d = b**2 - 4 * a * c
    if d == 0:
        return 'один'
    elif d < 0:
        return 'нет корней'
    else:
        return 'два'
```

Именованные функции

Определение

Определение функции позволяет **создать объект** функции. После определения к функции можно будет обратиться по заданному имени и получить результат ее работы.

В Python существует особый синтаксис определения именованных функций:

1. в начале - ключевое слово `def`
2. название функции

Существуют правила и рекомендации по именованию функций:

- **правила:**
 - название не может начинаться с числа
 - можно использовать нижнее подчеркивание `_`, цифры и буквы
 - **рекомендации:**
 - `snake_case`: в названии только буквы нижнего регистра, слова разделяются `_`
3. круглые скобки и, дополнительно, параметры внутри (*о них ниже*)
 4. двоеточие, переход на новую строку
 5. тело функции, выделенное отступом - набор инструкций, который "выполняет работу"

Код в Python организован в блоки и именно отступы дают понять, где у блоков начало и конец. Все тело функции должно располагаться минимум в одном отступе от начала строк.

```
def название_функции(позиционные, *доп_позиционные, именованные, **доп_именованные_параметры):
    инструкция 0
    инструкция 1
    ...
    инструкция N
```

Давайте напишем простейшую функцию, которая будет печатать две фразы:

```
# первым делом - def
# далее - имя функции "print_theme"
# после - круглые скобки, сейчас пустые. Потом : и переход на новую строку
def print_theme():

    # тело функции из двух вызовов print(), оба с единообразным отступом
    print('Лекция про функции!')
    print('Тело кончилось')
```

ВЫЗОВ

После определения функции появляется возможность к ней обращаться (вызывать). Делается это просто: указывается имя функции, круглые скобки и, дополнительно, аргументы в них.

```
# выше была определена print_theme, ее и вызовем
print_theme()
```

```
Лекция про функции!
Тело кончилось
```

Возвращаемое значение

Функция может не только производить какие-то действия, но возвращать наружу результат своей работы. Для этого используется ключевое слово `return`:

```
# определим функцию, которая принимает на вход x
def add_two(x):

    # переменная result - сумма x и 2
    result = x + 2

    # "наружу" возвращается полученное значение
    return result
```

На самом деле, любая функция что-то возвращает:

- указанное значение при наличии `return`
- `None`, когда `return` отсутствует

Возвращение результата означает, что его можно использовать вне функции, например, присвоить полученное значение переменной. Давайте посмотрим, что возвращает `print_theme()` и `add_two()`:

```
# Присвоим результат выполнения функции переменной и посмотрим, что в ней
from_print_theme = print_theme()
print(from_print_theme)
```

```
Лекция про функции!
Тело кончилось
None
```

Как видите, несмотря на отсутствие `return` в коде функции, она действительно возвращает `None`.

Теперь посмотрим на `add_two()`, где возвращаемое значение задано нами:

```
# add_two при вызове ожидает получить число: сложим 2 и 2
from_add_two = add_two(2)
print(from_add_two)
```

4

Пространства имен и области видимости

Чуть выше была определена `add_two()`, внутри которой инициализировалась переменная `result`. Можно ли получить значение, обратившись к ней вне функции? Нет:

```
# ошибочка!
# тут магическая конструкция try - except, которая "ловит" исключения и не
# дает коду перестать работать из-за ошибки
try:
    print(result)
except NameError as ne:
    print(ne)
```

```
name 'result' is not defined
```

Почему так происходит? Функции обладают своим пространством имен: переменные, которые определены внутри, не видны извне. Однако, функция может получать доступ к переменным, которые определены снаружи. Давайте опишем чуть более формально.

В Python объектам можно давать имя и по этому имени обращаться, эти названия организованы в пространства имен или неймспейсы (*от английского 'namespace'*). Инициализация переменной добавляет в неймспейс название объекта. Неймспейс - набор имен определенных на текущий момент объектов. Представьте себе `dict`, где ключом является строка, а значением - ссылка на объект. Область видимости определяет, переменные из каких пространств имен сейчас доступны. Разберем, как и где Python ищет объекты, к которым обращаются.

Неймспейсы и области видимости можно разделить на несколько типов:

- локальные (**Local**) - локальные переменные функции, данная область существует до тех пор, пока функция не завершит работу после вызова
- локальные пространства могут быть вложены друг в друга, в таком случае область уровня N (**Enclosing**) содержит "более глубокую" N + 1. Поиск имени, вызванного на уровне N, начинается с локального неймспейса этого уровня и продвигается "наружу", то есть на уровни выше
- глобальные (**Global**) - все объекты, определенные на уровне скрипта
- встроенные (**Built-in**) - неймспейс содержит все встроенные объекты (функции, исключения и т.д.), поиск в нем осуществляется последним



Если имя не было разрешено даже на уровне Built-in, будет выброшено исключение `NameError`.

```

z = 'name'

def foo():
    # в foo не определена своя z, но она успешно найдется в глобальной области
    print(z)
    a = 10
    b = -5

    def bar():
        # bar успешно напечатает a, при этом значение будет найдено в
        # неймспейсе foo
        # как думаете, где находится print?
        # в built-in, то есть самой "внешней" области
        print(a)

        # создадим переменную b внутри bar
        b = 20

        # будет напечатано 20, так как поиск увенчается успехом в локальном
        # неймспейсе
        print(b)

    bar()

    # если тут раскомментировать следующую строку, будет ошибка: foo не знает
    # ничего про внутренности bar
    # print(b)

foo()

```

```

name
10
20

```

💡 Tip

Хотим заметить, что нужно быть аккуратными при использовании вложенных функций и следить за тем, где и какие переменные определены и меняются. В рамках курса вообще не советуем так делать, чтобы не запутаться. Единственное исключение - декораторы, описанные ниже.

Параметры

Наша функция `add_two()` или, например, `type()` ожидают, что на вход будут переданы какие-то аргументы для успешной работы, а вот `print()` можно вызывать и с ними, и с пустыми скобками. В начале лекции был представлен скелет функции, сейчас разберем, что же находится в скобках.

Для начала немного формализма:

- при определении функции в скобках пишутся параметры, которые функция может принять
- при вызове функции в скобках указываются аргументы, которые задают значения параметров внутри функции

То есть имеется отображение: аргументы, с которыми вызывается функция -> значения параметров.

Переменные с названиями параметров могут быть использованы внутри тела функции, как будто их значения известны.

Позиционные параметры

Позиционные параметры выглядят как перечисленные внутри скобок названия переменных, используемых внутри функции:

```

# в данном случае есть два позиционных параметра
def foo(arg1, arg2):
    print(arg1, arg2)

```

Данный тип характеризуется следующим:

- позиционные параметры идут первыми после открытых скобок, все именованные строго после них
- важен порядок: отображение аргументов в параметры будет последовательным

- при вызове функции все аргументы, которые ожидает функция, должны быть переданы (откуда иначе Python возьмет значения? вот именно!)

Разберем пример, который суммирует два числа:

```
# в данном случае есть два позиционных параметра
def two_var_sum(var1, var2):

    # функция возвращает вычисленное значение суммы
    return abs(var1) + var2

# порядок важен!
print(two_var_sum(-1, 2), two_var_sum(2, -1))

# можно явно задавать переменные при вызове, в таком случае порядок не играет
# роли. Указывается название параметра и значение после =
print(two_var_sum(var2=2, var1=-1))
```

```
3 1
3
```

А что, если количество входных переменных очень большое или вы заранее не знаете, сколько аргументов будет передано при вызове? Например, вам нужно сложить не 2 числа, а 102? В Python есть специальный синтаксис со звездочками. После позиционных аргументов можно указать `list`, элементами которого станут неограниченное количество переданных позиционных аргументов.

Синтаксис: `имя_функции([поз0, ..., позN,] *поз_список): ... - []` в данном случае обозначают необязательность.

```
# тут позиционных аргументов нет (такое тоже может быть), поэтому сразу
# используется синтаксис со *: список, куда попадут все переданные позиционные
# аргументы, тут называется vars
def many_var_sum(*vars):

    # функция возвращает вычисленное значение суммы
    return sum(vars)

many_var_sum(1, 2, 3, 4, 5, 6)
```

```
21
```

💡 Tip

Совет: лучше передавать все в списках или векторах (о которых расскажут позже)

Именованные параметры

Данные параметры имеют значения по умолчанию.

Синтаксис: `имя_функции(все позиционные и *, им0=значение0, им1=значение1, ..., имK=значениеK, **им_словарь): ...`

характеризуются следующим:

- в определении идут строго после позиционных параметров
- в определении дано значение по умолчанию через `=` после имени
- при вызове необязательно передавать - тогда будет использовано значение по умолчанию

```

# тут есть один позиционный параметр и один именованный
def hello(name, phrase='Привет'):

    print(f'{phrase}, {name}')

# передавать значения аргументов можно как в исходном порядке, тогда
# параметру будет присвоено соответствующее значение
hello('Саша', 'Приветствую')

# так и указывая названия параметров
hello(phrase='Здорова', name='Игорь')

# если не указать значение именованного параметра, используется дефолтное
hello('Вася')

```

Приветствую, Саша
Здорова, Игорь
Привет, Вася

Аналогично позиционным аргументам, если необходимо передать множество именованных параметров, используется синтаксис со звездочками. В данном случае все переданные именованные аргументы, если не определены явно, попадут в `dict`, указанный после **:

```

def congrats(today, everyone=False, **names):
    """
    Функция может поздравлять людей
    Args:
        today (str): сегодняшняя дата
        everyone(bool): флаг, нужно ли поздравить всех
        names (dict): отображение имя: дата
    """

    if everyone:
        print('Happy Birthday!')

    else:
        for name, date in names.items():
            if date[-5:] == today[-5:]:
                print(f'Happy Birthday, {name}')

```

congrats('2021-09-17', Paul='2001-03-08', Lena='1997-01-31', Mark='1997-09-17')

Happy Birthday, Mark

Что это за комментарий в кавычках внутри функции? Это один из общепринятых способов написания [docstring](#) - описания деталей работы функции.

АНОНИМНЫЕ ФУНКЦИИ

Функции, определенные при помощи `def`, имеют название, по которому можно обратиться, но также существуют и анонимные или неименованные функции.

Такие функции могут быть определены при помощи оператора `lambda`. Он позволяет задать входные и выходные данные. Вы можете попробовать самостоятельно почитать [документацию](#).

```
# синтаксис анонимных функций простой
lambda [арг0, ..., аргN]: выражение
```

В определении `выражение` собой то, что будет возвращено из анонимной функции:

```

lambda x: abs(x)

lambda num, div=2: 'нет' if num % div else 'да'

```

<function __main__.<lambda>(num, div=2)>

В примере выше `lambda`-функции были только созданы в моменте, но не вызваны. Можно сказать, что мы их определили (как с `def`), но не воспользовались ими, по сути, сразу потеряли.

Для удобства можно присвоить переменной объект функции и по этому имени к ней обращаться:

```
# сохранили в check_div функцию
check_div = lambda num, div=2: 'нет' if num % div else 'да'

# вызов совершенно обычный
print(check_div(3), check_div(5, 5))

# если без сохранения в переменную:
print(lambda x: abs(x))(-120)
```

```
нет да
120
```

Данные функции можно применять, когда нужна одноразовая функция или лень писать лишнюю строку с `def`.

Декораторы

Полезный концепт Python - декоратор. Это функция, которая принимает на вход другую функцию и возвращает функцию. Звучит непонятно на первый взгляд. На деле декораторы помогают расширять возможности функций не меняя их кода.

Давайте сразу разберем пример:

```
import time

# на вход декоратор принимает параметр func - оборачиваемую функцию
def time_decorator(func):

    # внутри определена функция, которая "заменит" переданную
    # wrappred будет засекать, за сколько выполняется переданная функция
    def wrapped(*args):
        start = time.time()

        # почему func нет среди аргументов wrappred?
        # все аргументы, которые примет wrapped, тут передаются в func
        result = func(*args)

        end = time.time()
        print(f'Прошло {end - start} секунд')

        return result

    # возвращает декоратор тоже функцию. обратите внимание, что
    # возвращаемое значение именно объект function: после имени нет круглых
    # скобок. если бы они были, возвращался бы результат выполнения wrappred,
    # так как wrappred() - вызов функции
    return wrapped
```

Выше определена функция `many_var_sum()`, давайте засекать, сколько она работает:

```
# сохраним обернутую функцию (помним, декоратор возвращает функцию) в
# переменную, обратите внимание, что на вход декоратору передается также
# объект самой функции, а не результат ее работы
many_var_sum = time_decorator(many_var_sum)

# поведение функции поменялось, а код - нет
summed = many_var_sum(10, 0, -120, 333)
```

```
Прошло 1.1920928955078125e-06 секунд
```

Отлично, но можно добавить так называемый **синтаксический сахар**: вместо присваивания значения над определением функции можно указать специальный символ `@`, после которого указывается название декоратора:

```
# теперь, при вызове, stupid_power сразу будет обернута!
@time_decorator
def stupid_power(x, power=5):
    result = 1
    for p in range(power):
        result *= x
    return result

powered = stupid_power(10)
```

```
Прошло 3.337860107421875e-06 секунд
```

Таким образом, использования синтаксического сахара с `@` и указанием имени декоратора над функцией аналогично вызову `stupid_power = time_decorator(stupid_power)`.

Что мы узнали из лекции

- Что такое функции и зачем их применять
- Как определить функцию (инструкция `def`)

```
def название_функции(параметры):  
    тело функции  
<return возвращаемые_значения>
```

- Отличия позиционных параметров от именованных:
 - порядок указания - сначала позиционные, *потом* именованные
 - значения по умолчанию у именованных позволяют не указывать их при вызове
- Синтаксис со звездочками для получения заранее неизвестного числа позиционных (`*some_list`) и именованных (`**some_dict`) аргументов
 - в `*some_list` стоит **одна** `*`: эта конструкция для получения неограниченного количества позиционных/неименованных аргументов
 - в `**some_dict` **две** `*`: все именованные аргументы, явно не указанные среди параметров, попадут туда
- Что может возвращать функция при помощи `return`
 - `return` обозначает выход из функции и передачу "наружу" результата работы
 - в `return` можно перечислять несколько возвращаемых значений через запятую
- Что такое декораторы и как они работают, щепотку синтаксического сахара с `@`
 - использование декоратора эквивалентно сохранению результата вызова функции-декоратора с аргументом в виде оборачиваемой функции `stupid_power = time_decorator(stupid_power)`

Объекты, классы и методы

Описание лекции

В этой лекции мы расскажем про:

- основы ООП: что это и зачем?
- инстанс класса
- на что указывает `self`
- методы классов в уже знакомых примерах
- атрибуты классов

ООП: Объектно-Ориентированное Программирование

Что же такое объектно-ориентированное программирование? Судя по названию, ключевую роль здесь играют объекты, на которые ориентируется дальнейший процесс программирования. Если мы взглянем на реальный мир, то для нас он представит в виде множества объектов, обладающих определенными свойствами, взаимодействующих между собой и вследствие этого изменяющимися. Эта привычная для взгляда человека картина мира была перенесена в программирование. Python - это объектно-ориентированный язык программирования, и **все в нем является объектами**.

Объект - это набор данных и инструкций в памяти компьютера, в том числе:

1. Тип объекта;
2. Данные, формирующие объект (контент внутри него);
3. Методы.



Fig. 35 Так можно проиллюстрировать соотношение “класс - объекты”. Сверху указаны общие прототипы (прямоугольник, звезда, треугольник), снизу - уникальные экземпляры, со своим цветом и деталями.

Как определить тип мы уже знаем. И с объектами на самом деле мы знакомы с самого начала - ведь даже целочисленная переменная `5`, имеющая `type(5) = int`, представляет собой объект. Напомним, что говорилось в курсе ранее: **методы - это те же функции, только существующие не сами по себе, а являющиеся частью класса.**

```
example_list = list(["что-то", "с"])
print(f"Метод append имеет тип {example_list.append}")

# вот так можно вызвать метод, изменяя сам экземпляр класса
example_list.append("чем-то")

# а вот так можно обратиться к атрибуту конкретного объекта
attr_value = example_list.__doc__
print(f"Значение атрибута: {attr_value}")
```

Метод `append` имеет тип `<built-in method append of list object at 0x7f2022e67640>`
Значение атрибута: `Built-in mutable sequence`.

If no argument is given, the constructor creates a new empty list.
The argument must be an iterable if specified.

`list` - это класс, и в первой строчке кода выше создается **объект, или экземпляр класса `example_list`**. Он хранит свой набор данных (контент). Метод `append` общий для всех списков, однако меняет только конкретный экземпляр. Обратите внимание: этот метод ничего не возвращает, он меняет сам объект и его контент. Не все методы работают схожим образом, и придется запоминать принцип работы самых распространенных методов. К примеру, сложение чисел или строк внутри Python также осуществляется с помощью методов, однако это не меняет саму переменную, как было рассказано в более ранних лекциях:

```
# создадим два объекта класса int
a = 3
b = int(4)

# неявно вызовем метод сложения
a + b
# проверим, что содержимое объектов не изменилось
print(f"{a} , {b}")
```

`a = 3, b = 4`

Конкретно в данном случае логика работы обусловлена свойствами `mutable/immutable` объектов, о чем также упоминалось в одной из предшествующих лекций. `int` неизменяем, и потому при вызове метода сложения возвращает новый объект - ведь самого себя изменить не получится. `list` же может меняться, и `append` меняет содержимое `inplace`.



Fig. 36 Другой житейский пример - по одному чертежу (классу) можно построить разные (но похожие!) дома, в которых разные семьи, убранство, разные обои.

Определение классов

Давайте попробуем создать свой первый класс:

```
# определим классы
class List:
    def __init__(self, initial_content = None):
        # указывать в виде значения по умолчанию mutable-объекты неправильно
        if initial_content is None:
            initial_content = []
        self.content = initial_content

    def append(self, new_element):
        # изменение производится inplace, с заменой собственного контента
        self.content = self.content + [new_element]

class Int:
    def __init__(self, value):
        self.value = value

    def add(self, second_value):
        # изменение не производится, создается новый объект и он же возвращается
        return Int(self.value + second_value.value)

# воспроизведем примеры кода выше для List
example_list = List(["что-то", "с"])
print(f"{example_list.content = }")

example_list.append("чем-то")
print(f"{example_list.content = }")
print(f"type(example_list) = {type(example_list)}")

# и для Int
a = Int(3)
b = Int(4)

a.add(b)
print(f"После простого сложения: {a.value = }, {b.value = }")

c = a.add(b)
print(f"После записи в переменную результата: {c.value = }")
```

```
example_list.content = ['что-то', 'с']
example_list.content = ['что-то', 'с', 'чем-то']
type(example_list) = <class '__main__.List'>
После простого сложения: a.value = 3, b.value = 4
После записи в переменную результата: c.value = 7
```

Что важно заметить и отметить в этом примере:

1. Как объявить класс? С помощью ключевого слова `class`, указания названия класса (таким образом создается новый тип данных) и перечисления методов:

```
class ИмяКласса:
    <код_тела_класса>
```

- Обязательно определить метод `__init__`. Это метод, вызываемый при создании нового объекта класса и отвечающий за определение его контента. В нем через запятую, как и в обычной функции, перечислить необходимые переменные и, по возможности, их значения по умолчанию;
- Каждый метод должен первым аргументом принимать **экземпляр объекта** (по соглашению он называется `self`). При этом при вызове метода Python автоматически подставит сам объект, делать это руками не нужно. По сути под капотом происходит следующее:

```
# вызываем метод класса - не объекта! - и передаем первым аргументом объект,
# который при обычных обстоятельствах будет записан в self
result = Int.add(Int(3), Int(4))
print(f"{result.value}")
```

7

- Атрибутом называется переменная, хранящаяся в экземпляре класса;**
- В рамках методов атрибуты объекта доступны с использованием ключевого слова `self`. К примеру,
`self.data;`
- Вне методов атрибуты доступны у экземпляров через точку: `some_object.attr_name`;
- Для создания объекта класса нужно указать имя класса (типа) и передать аргументы: `Int(3)` - как раз оно;
- Методы как могут возвращать что-либо (в том числе новый объект того же класса), так и ничего не возвращать.

В заключение еще одна картинка, иллюстрирующая суть класса, экземпляра, атрибутов и методов:



Fig. 37 Who's that pokemon?

Пример использования для хранения состояния

В курсе будет встречаться подобный паттерн, поэтому хочется остановиться поподробнее. Допустим, есть некоторая среда, заданная состоянием, и есть объекты, существующие в этой среде, как-то взаимодействующие между собой (и со средой). Есть набор функций, который осуществляет эти взаимодействия. Тогда удобно ввести некоторый объект `system` класса `System`, который хранит эту информацию, и передавать его как параметр в методы расчета чего-либо (никакой конкретной логики тут не заложено, не пытайтесь понять, за что отвечает `alpha` и прочие переменные - пример абстрактный):

```

# класс системы, с ее параметрами и состоянием
class System:
    def __init__(self, param_vector, alpha, gamma, multiplier = 2.0, energy = 0.0):
        self.param_vector = param_vector
        self.alpha = alpha
        self.gamma = gamma
        self.multiplier = multiplier
        self.energy = energy

        # начальное состояние
        self.system_state = [
            (alpha - gamma) * multiplier * param
            for param
            in self.param_vector
        ]

    def first_action(self, object):
        # взаимодействие с одним объектом: увеличение энергии, изменение состояния
        # системы
        self.gamma -= object.energy
        self.energy += object.energy
        self.system_state = [param - self.alpha * self.energy for param in
        self.system_state]

    def second_action(self, list_of_object):
        # взаимодействие с несколькими объектами - увеличение их энергии
        for obj in list_of_object:
            obj.energy += self.gamma * self.multiplier

    class Object:
        def __init__(self, energy = 10.0):
            # объект хранит только один атрибут
            self.energy = energy

    # симуляция одного цикла в системе
    def one_system_cycle(system: System, objects = None):
        if objects is not None:
            system.second_action(objects)
            # убираем один последний объект
            objects = objects[:-1]
        else:
            objects = [Object(val) for val in range(5)]
            for obj in objects:
                system.first_action(obj)
        return objects

    params = [1.0, 3.0, 4.15, 0.0]
    alpha = 5
    gamma = 0.18
    system = System(params, alpha, gamma, energy = 10)

    # изначально объектов нет, они появятся после первого цикла
    objects = None

    for cycle_num in range(1, 4):
        print(f'Состояние системы на итерации {cycle_num}: {system.system_state}.',
        end="\t")
        objects = one_system_cycle(system, objects)
        print(f'Всего объектов: {len(objects)}', end="\t")

        total_object_energy = sum([obj.energy for obj in objects])
        print(f'Суммарная энергия объектов: {total_object_energy}')

```

Состояние системы на итерации 1: [9.64, 28.92, 40.0060000000001, 0.0]. Всего объектов: 5 Суммарная энергия объектов: 10
 Состояние системы на итерации 2: [-340.36, -321.08, -309.994, -350.0]. Всего объектов: 4 Суммарная энергия объектов: -72.56
 Состояние системы на итерации 3: [-340.36, -321.08, -309.994, -350.0]. Всего объектов: 3 Суммарная энергия объектов: -114.84

Обратите внимание на то, что функция `one_system_cycle` не возвращает `system` - ведь мы прямо во время итераций меняем ее состояние, и в последующие моменты времени эти изменения сохраняются, то есть состояние остается. Таким образом, эту систему (среду) - а именно объект, экземпляр класса `System` - как параметр можно передавать в десятки функций, и каждая из них будет видеть актуальное состояние.

Что мы узнали из лекции

- Объекты одного класса отличаются между собой атрибутами, которые придают уникальности;
- Методы классов позволяют описывать логику взаимодействия, **обновлять состояние**;
- Для объявления класса нужно определить **как минимум один метод `__init__`**;

- Классы без методов можно использовать как **удобное хранилище разнородной информации** о схожих объектах (описать разных студентов и их оценки).

Это далеко не все, что можно написать по теме классов и объектов, однако это тот необходимый минимум, что потребуется для прохождения курса. Больше примеров и деталей можно найти, например, по ссылкам: [пример с несколькими взаимодействующими объектами, что такое наследование и зачем оно нужно](#).

Финальная лекция обо всем и ни о чём

Описание лекции

В этой лекции будет:

- рассказ о типовых ошибках при прохождении курса;
- разбор примеров кода из курса;
- дополнение для самых любознательных о **Python**, ML и вообще.

Заметки по прохождению курса

Хоть в пройденных лекциях по **Python** и содержится большое количество материала, проиллюстрированного примерами, с указанием на дополнительные темы для самостоятельного изучения – это все еще малая доля того, с чем вам придется столкнуться, если захочется продолжить развитие в теме ML/QML. Да, это грустно, но такова правда жизни. А еще вы столкнетесь с кучей проблем и ошибок во время прохождения самого курса. Если вы **попытались разобраться** с чем-либо, но у вас не получилось – приходите к нам в сообщество в Slack, задавайте вопросы. Не забывайте и о взаимопомощи другим студентам – если материал или тема дались вам легко, то кому-то они могут показаться адски сложными.

Помните, что абсолютное большинство возникающих ошибок в **Python** неуникально, и вы не первый, кто наступил на эти грабли. Не стесняйтесь пользоваться поисковиками для нахождения решений. Но перед этим обязательно прочитайте текст ошибки, попробуйте осознать, о чем **Python** хочет вам сказать – в этом языке программирования одни из самых понятных и доступных подсказок и указателей на проблемы. К примеру, **IndexError: list index out of range** явно говорит о том, что во время обращения к i-тому объекту некоторого списка (который, кстати, будет указан в так называемом трейсбеке, от англ. traceback) возник выход за границы **list**. Возьмем для примера список **sample_list** чисел длины 5. Что произойдет, если обратиться к седьмому объекту **sample_list[7]**?

```

Ввод [36]: sample_list = [1, 2, 3, 4, 5]

Ввод [39]: print('что-то')
print('И еще что-то')
example_str = 'а тут я ' + ' просто складываю строки'
i_umnozhau_chisla = 3*6
result = sample_list[7]
# а после этого еще оставлю пару строк
b = 'с какими то караулами'

что-то
И еще что-то

-----
IndexError: list index out of range
Traceback (most recent call last)
<ipython-input-39-567b10e20d31> in <module>
      3 example_str = 'а тут я ' + ' просто складываю строки'
      4 i_umnozhau_chisla = 3*6
----> 5 result = sample_list[7]
      6 # а после этого еще оставлю пару строк
      7 b = 'с какими то караулами'

IndexError: list index out of range

```

Fig. 38 Получим ошибку

Обратите внимание на стрелочку слева, указывающую на строку номер 5 (кстати, в Jupyter Notebook для каждой отдельной ячейки можно включить нумерацию строк для удобства навигации с помощью клавиши **l**). Это место возникновения ошибки, та строка, на которой **Python** не смог выполнить инструкцию. Как можно решить проблему? Для начала **методом пристального взгляда** поймите, написано ли в коде то, что вы хотели – очень часто код, к примеру, копируется (хотя нужно избегать этого с помощью написания функций), но переменные в нем не меняются в силу невнимательности. Если все переменные в строке те же, что и должны быть, и ошибка неясна, то следует заняться [отладкой кода](#).

Самый простой способ отладки в Jupyter – это расстановка **print**'ов до проблемной строки. Да-да, так кустарно и грубо. Есть более продвинутые технологии (например, [отладчики](#) в IDE, то есть средах разработки), однако они выходят за рамки данного курса. Очевидный совет – нужно размещать такой **print**, который помогает проверить

какую-либо **гипотезу** о происхождении ошибки. Предложение таких гипотез – это отдельный навык, который приходит с опытом, так что не переживайте, что поначалу получается плохо “угадывать” и “понимать” ошибки.

В случае обращения по неверному индексу (`list index out of range`) логично предположить, что индекс оказался больше максимального возможного значения. Давайте выведем длину списка и индекс, по которому хотим обратиться:

```
Ввод [36]: sample_list = [1, 2, 3, 4, 5]

Ввод [40]: print('что-то')
print('И еще что-то')
example_str = 'а тут я ' + ' просто складываю строки'
i_umnozhau_chisla = 3*6
target_index = 7
print(f'ОТЛАДКА: {len(sample_list) = }, {target_index = }', )
result = sample_list[target_index]
# а после этого еще оставляю пару строк
b = 'с какими то каракулами'

что-то
И еще что-то
ОТЛАДКА: len(sample_list) = 5, target_index = 7

-----
IndexError: Traceback (most recent call last)
<ipython-input-40-c3b40d282224> in <module>
      5 target_index = 7
      6 print(f'ОТЛАДКА: {len(sample_list) = }, {target_index = }', )
----> 7 result = sample_list[target_index]
      8 # а после этого еще оставляю пару строк
      9 b = 'с какими то каракулами'

IndexError: list index out of range
```

Fig. 39 Строчку с вашим отладочным выводом легко потерять - для удобства можно добавлять какие-то символы или текст.

Исходя из аргументов выше становится очевидно, в чем именно проблема. Конечно, рассмотренный случай крайне тривиален, однако он показывает базовые приемы поиска причин ошибок. Отметим также такой полезный прием как разбитие проблемной строки кода на подстроки, то есть разнесение кода на отдельные изолированные куски. В данном примере в отдельную переменную был записан индекс. Однако попробуйте угадать, в каком месте происходит ошибка в коде ниже:

```
Ввод [41]: def first_f(inp):
    return inp * 3

def second_f(inp):
    return inp - 2

def third_f(inp):
    return inp + 'abracadabra'

print(first_f(third_f(second_f(12))))


-----
TypeError: Traceback (most recent call last)
<ipython-input-41-16694fbb215b> in <module>
      8     return inp + 'abracadabra'
      9
----> 10 print(first_f(third_f(second_f(12))))
<ipython-input-41-16694fbb215b> in third_f(inp)
      6
      7 def third_f(inp):
----> 8     return inp + 'abracadabra'
      9
     10 print(first_f(third_f(second_f(12)))))

TypeError: unsupported operand type(s) for +: 'int' and 'str'
```

Fig. 40 Трейсбек ошибки предательски указывает на функцию, при вызове которой возникла ошибка - `third_f`

Это уже более сложный пример, когда трейсбек состоит из двух блоков. Указатель на ошибку как бы “проваливается” внутрь вызываемой функции и показывает, что не так. В данном случае ошибка `TypeError: unsupported operand type(s) for +: 'int' and 'str'` указывает на несовместимость типов операндов в операции сложения. Слева `int`, справа `str` – ну и как же их суммировать?

Однако исходная строчка, порождающая ошибку, крайне сложна. Мы вызываем функцию (`print`) от функции (`first_f`) от функции... Будь логика в коде чуть сложнее – легко можно было бы запутаться, что именно породило ошибку (главное обращайте внимание на заголовки трейсбека, указывающие на названия методов и функций). С разбиением все проще:

```

Ввод [42]: def first_f(inp):
    return inp * 3

def second_f(inp):
    return inp - 2

def third_f(inp):
    return inp + 'abracadabra'

tmp1 = second_f(12)
tmp2 = third_f(tmp1)
tmp3 = first_f(tmp2)

print(tmp3)

-----
TypeError                                 Traceback (most recent call last)
<ipython-input-42-5e2fb8ada7f> in <module>
      9
     10 tmp1 = second_f(12)
--> 11 tmp2 = third_f(tmp1)
     12 tmp3 = first_f(tmp2)
     13

<ipython-input-42-5e2fb8ada7f> in third_f(inp)
      6
      7 def third_f(inp):
-->  8     return inp + 'abracadabra'
      9
     10 tmp1 = second_f(12)

TypeError: unsupported operand type(s) for +: 'int' and 'str'

```

Fig. 41 Вуали! Указатель в виде стрелочки прямо рапортует о том, что проблема в вычислении `third_f`. Дальше можно было бы добавить `print(tmp1)` для того, чтобы увидеть входные параметры и попытаться понять, в чем же проблема.

На этом остановимся, блок обязательно будет дополняться по мере возникновения типовых ошибок и способов их поимки. Главное помните, что можно распечатать и аргументы функции/метода, и атрибуты класса (в том числе скрытые/промежуточные состояния квантовой среды), и результаты расчетов.

Примеры кода, не разобранные в лекциях

Ранее весь код, который упоминался в лекциях (за вычетом примера из первой) был написан на чистом Python без сторонних библиотек, которые мы так нахваливали. В курсе же, разумеется, они встретятся. Полезно поговорить о двух вещах - об **импортах** библиотек и об их алиасах.

Для того, чтобы использовать какой-либо объект из сторонней библиотеки, будь то функция, класс или даже словарь (например, с физическими константами), необходимо знать, в какой библиотеке и по какому пути он лежит. Все это описано в документации, а также в примерах кода в курсе. Не переживайте, если они плохо запоминаются - чаще всего их копируют, и со временем в голове откладывается паттерн. И все же для того, чтобы получить доступ к библиотеке, нужно сделать следующее – рассмотрим на примере `numpy`, библиотеки для алгебраических расчетов. Сначала следует ключевое слово `import`, затем через пробел название пакета. Опционально через `as` можно добавить алиас - название, под которым можно будет обращаться в коде для доступа к функциям. У многих пакетов есть общепринятые алиасы, с которыми все знакомы. Для `numpy` это `np`.

Итак,

```

import numpy as np

np_example_array = np.array([1,2,3])
print(f'{np_example_array.shape = }')

abs_res = np.abs(-2)
print(f'{abs_res = }')

```

```

np_example_array.shape = (3,)
abs_res = 2

```

По сути обращение схоже с объявлением объекта класса или же каким то методом - ведь можно импортировать не целый пакет, а его часть!

```

from numpy import zeros

zeros_matrix = zeros(shape=(1, 3))
print(f'{zeros_matrix = }')

```

```

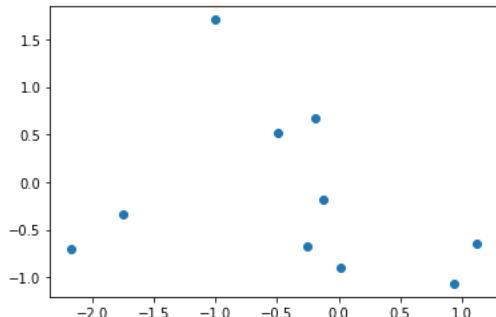
zeros_matrix = array([[0., 0., 0.]])

```

Здесь `zeros` - **функция**, формирующая матрицу из нулей заданного размера.

Или, к примеру, для отрисовки визуализаций часто используют следующую библиотеку (мы уже с ней сталкивались):

```
import matplotlib.pyplot as plt  
plt.scatter(np.random.randn(10), np.random.randn(10));  
plt.show();
```



Tip

В коде выше в последних двух строчках стоит `;`. Это необязательный символ, означающий конец команды. В большинстве других языков программирования необходимо всегда в конце строки оставлять этот символ, однако в питоне сигналом о конце инструкции служит перенос строки. Почему же здесь они поставлены? Во-первых, чтобы показать, что они означают – чтобы вы не переживали, когда встретите подобное. Во-вторых, `plt.show()` возвращает объект после вызова метода. Jupyter же по умолчанию выводит последнее возвращенное значение. Попробуйте убрать точку с запятой и посмотрите в своем ноутбуке, что выйдет. Таким образом, в данном случае окончание команды `;` говорит, что возвращаемая переменная никуда не будет записана и будет удалена – поэтому при ее наличии подпись объекта не возвращается.

В коде сначала происходит импорт, затем вызывается отрисовка точек (см. [документацию тут](#)), где координаты `X` и `Y` получаются случайно из нормального распределения от `-1` до `1`. Но сам по себе вызов `scatter()` лишь добавляет точки на график, но не отрисовывает его – за это отвечает вызов `plt.show()`. Стоит заметить, что `plt` хранит в себе “полотно”, на котором рисует, и результат добавления точек через `scatter` не нужно записывать в какую-либо переменную – все это хранится во внутреннем состоянии самой библиотеки. Помните урок про классы и их объекты, задающие состояние системы? Здесь ровно то же самое! Более того, аналогичный функционал будет у библиотеки для квантовых вычислений, с которой вы познакомитесь далее по курсу (под алиасом `qml`). Теперь, надеемся, картинка начинает складываться!

Блок дополнительных ссылок

Ниже представлен набор ссылок, которые предлагаются просмотреть для расширения кругозора и лучшего понимания материала. Это бонусный материал, необязательный к изучению. Но если вас захватила идея получше познакомиться со своим инструментом, то почему нет?

1. [Курс по Python](#) в виде примеров кода с очевидными и не очень механиками языка. Хорошо подойдет как дополнение к примерам в курсе;
2. [realpython](#) – сайт с большим количеством разнообразных уроков по Python на английском;
3. [Огромный список](#) ссылок на разные статьи по всем темам - от основ Python и до продвинутых приемов в разработке;
4. [И другой](#) – с набором ссылок на код разных программ. Игры, видеопроигрыватели и прочее - все тут;
5. [Простой отладчик](#) на случай, если начали путаться в своем коде, а продвинутые инструменты кажутся очень сложными и громоздкими.

Tip

Как ставить новые библиотеки в Python – вот хотя бы тот же отладчик из пункта 5? Вспоминайте совет – гугл ваш лучший друг. “how to install python library” (да, прямо дословно library = библиотека) – и вы на коне!

Вместо заключения

Что ж, вот и подошел к концу первый блок курса. Его можно назвать самым сложным – нет, не с точки зрения материала и его научности, но по количеству усилий, которые необходимо приложить, чтобы не сбиться с пути и продолжить обучение несмотря на трудности. Именно тут большая часть людей машет рукой и закрывает курс. Если ты, дорогой студент, читаешь это, то мы безумно рады за твои достижения и подвиги, и желаем дальнейших успехов в прохождении курса!



О блоке “Линейная алгебра и NumPy”

Блок задуман для обсуждения основных понятий и идей линейной алгебры и представляет собой краткое пособие. Данный блок **не является** полноценным учебником по линейной алгебре. Для более детального погружения в тему рекомендуем учебник Тыртышникова Е. Е. [«Матричный анализ и линейная алгебра»](#).

В рамках блока мы:

- обсудим основные понятия и идеи, такие как **векторы и матрицы**;
- познакомимся с инструментами для работы с векторами и матрицами;
- попробуем что-нибудь посчитать!

Обозначения:

- $\backslash(a)$ - число (скаляр)
- $\backslash(\mathbf{a})$ - вектор
- $\backslash(\mathbf{A})$ - матрица

Полезные материалы:

- MathProfi: [Аналитическая геометрия](#)
- 3Blue1Brown: [Essence of linear algebra](#)

Векторы

Определение

Вектор – это термин, который имеет несколько различных интерпретаций: математическую, геометрическую, физическую и др. Точный смысл данного термина зависит от контекста.

Формально вектор определяется как элемент векторного пространства – множества, на котором определены операции **сложения и умножения вектора на число (скаляр)**, которые должны удовлетворять [8 аксиомам](#).

Для простоты понимания рассмотрим знакомую нам со школьных времен прямоугольную (Декартову) систему координат на плоскости – две перпендикулярные друг другу оси $\backslash(x)$ и $\backslash(y)$, выбранные на них единичные векторы (орты) $\backslash(\mathbf{e}_1)$, $\backslash(\mathbf{e}_2)$ и начало координат.

Вектор $\backslash(\mathbf{a})$ в такой системе координат можно записать следующим образом: $\backslash(\mathbf{a} = \mathbf{a}_1 \mathbf{e}_1 + \mathbf{a}_2 \mathbf{e}_2 = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix})$

Замечание 1: Координаты вектора не определяют однозначно его положение на плоскости, а лишь положение конца вектора относительно его начала. Например, вектор $\backslash(\begin{pmatrix} 3 \\ 4 \end{pmatrix})$ может быть отложен как из точки начала координат $\backslash(\begin{pmatrix} 0 \\ 0 \end{pmatrix})$ с концом в точке $\backslash(\begin{pmatrix} 3 \\ 4 \end{pmatrix})$,

`\end{pmatrix}), так и из произвольной точки, например, \begin{pmatrix} 1 , 1 \end{pmatrix} с концом в точке \begin{pmatrix} 4 , 5 \end{pmatrix}. Оба этих вектора соответствуют 3 ортам на оси \(x) и 4 на оси \(y). Обычно, если не сказано иное, предполагается, что вектор отложен из начала координат.`

Замечание 2: Вектор можно представить либо как вектор-столбец `\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}`, либо как вектор-строку `\begin{pmatrix} a_1 & a_2 \end{pmatrix}`. Здесь и далее под вектором будем подразумевать вектор-столбец, если не сказано иного.

Понятие вектора на плоскости можно обобщить на 3-мерное пространство, и, в общем случае, на (n) -мерное пространство (которое уже не получится визуализировать):

$$\begin{aligned} \begin{aligned} \mathbf{a} = a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + \dots + a_n \mathbf{e}_n &= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \\ &\quad \text{или} \\ &= a_1 \mathbf{e}_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 \mathbf{e}_2 + \dots + a_n \mathbf{e}_n \mathbf{e}_n \end{aligned} \end{aligned}$$

Операции с векторами

Как было сказано ранее, в формальном определении существуют две основные операции над векторами:

- сложение:

$$\begin{aligned} \begin{aligned} \mathbf{a} + \mathbf{b} &= (a_1 + b_1) \mathbf{e}_1 + (a_2 + b_2) \mathbf{e}_2 + \dots + (a_n + b_n) \mathbf{e}_n \\ &= \begin{pmatrix} a_1 & a_2 & \vdots & a_n \end{pmatrix} + \begin{pmatrix} b_1 & b_2 & \vdots & b_n \end{pmatrix} \\ &= \begin{pmatrix} a_1 + b_1 & a_2 + b_2 & \vdots & a_n + b_n \end{pmatrix} \end{aligned} \end{aligned}$$

- умножения вектора на число (скаляр):

$$\begin{aligned} \begin{aligned} \lambda \mathbf{a} &= \lambda a_1 \mathbf{e}_1 + \lambda a_2 \mathbf{e}_2 + \dots + \lambda a_n \mathbf{e}_n \\ &= \lambda \begin{pmatrix} a_1 & a_2 & \vdots & a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 & \lambda a_2 & \vdots & \lambda a_n \end{pmatrix} \end{aligned} \end{aligned}$$

Операцию умножения вектора на число можно интерпретировать геометрически, как сжатие / растяжение вектора.

Используя эти две операции, мы можем считать **линейные комбинации векторов**: $\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \dots + \alpha_n \mathbf{a}_n$, где $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ – векторы, а $(\alpha_1, \alpha_2, \dots, \alpha_n)$ – числа.

Норма (длина) вектора

В линейной алгебре для обобщения понятия длины вектора используется термин **норма**. Можно сказать, что понятия длины и нормы эквивалентны.

Формально норма определяется как функционал в векторном пространстве, удовлетворяющий [3 аксиомам](#) и отображающий элементы этого пространства (векторы) в множество неотрицательных вещественных чисел.

Данному определению нормы удовлетворяет множество функционалов, но мы будем рассматривать наиболее часто используемый – **Евклидову норму**.

Для простоты понимания рассмотрим вектор на плоскости. С геометрической точки зрения он представляет собой направленный отрезок. Направленность вектора никак не влияет на его длину, поэтому, при расчете длины, мы можем работать с ним как с отрезком на плоскости и посчитать длину по школьным формулам. Заметим тот факт, что координаты вектора соответствуют числам, умноженным на орты координатных осей, поэтому формула длины вектора выглядит следующим образом: $\sqrt{\|\mathbf{a}\|}$. Соответственно, в общем случае формула выглядит следующим образом: $\sqrt{\|\mathbf{a}\|^2} = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$.

Скалярное произведение

Одной из самых распространенных операций над двумя векторами является так называемое **скалярное произведение**, результатом которого является число (скаляр) – отсюда и название операции.

Замечание: Кроме скалярного произведения существует также [векторное произведение](#) над парой векторов, результатом которого являются вектор. Также существует [смешанное произведение](#) над тройкой векторов, результатом которого является число. Данные операции в рамках курса рассматриваться не будут.

Скалярное произведение используется в определении длины векторов и угла между ними. Данная операция имеет два определения:

- алгебраическое: $\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$

- геометрическое: $\|\mathbf{a}\| \cdot \|\mathbf{b}\| = \sqrt{\mathbf{a} \cdot \mathbf{b}} = \sqrt{\cos(\theta)}$, где θ – угол между векторами \mathbf{a} и \mathbf{b} .

Используя оба эти определения можно вывести формулу для расчета косинуса угла между векторами:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} = \frac{\mathbf{a}_1 b_1 + \mathbf{a}_2 b_2 + \dots + \mathbf{a}_n b_n}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}}$$

С помощью данной формулы можно прийти к одному из главных свойств скалярного произведения, которое заключается в том, что **два вектора перпендикулярны тогда и только тогда, как их скалярное произведение равно 0**: $\mathbf{a} \cdot \mathbf{b} = 0 \iff \cos(\theta) = 0$

Скалярное произведение можно использовать для расчета нормы вектора следующим образом: $\|\mathbf{a}\| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$

Линейная независимость

Одним из основополагающих понятий линейной алгебры является **линейная зависимость/независимость**.

Для определения данного понятия рассмотрим набор из нескольких векторов. Набор векторов является **линейно зависимым**, если существует такая ненулевая линейная комбинация векторов данного набора (как минимум один элемент данной комбинации не равен 0) равная нулевому вектору (вектор, состоящий только из 0):

$$\begin{aligned} \lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + \dots + \lambda_m \mathbf{a}_m &= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ \exists \lambda_1 \neq 0 \vee \lambda_2 \neq 0 \vee \dots \vee \lambda_m \neq 0 \end{aligned}$$

Если набор векторов не является линейно зависимым, то есть не существует ненулевой линейной комбинации векторов данного набора равной нулевому вектору, то такой набор векторов называется **линейно независимым**.

Пример:

- линейно независимый набор векторов (только нулевые коэффициенты линейно комбинации приводят к нулевому вектору):

$$\begin{aligned} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{aligned}$$

- линейно зависимый набор векторов (существуют ненулевые коэффициенты линейно комбинации, которые приводят к нулевому вектору):

$$\begin{aligned} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{aligned}$$

Из определения линейной зависимости можно вывести следующее свойство: Набор векторов линейно зависим тогда и только тогда, когда один из элементов этого набора может быть выражен через оставшиеся.

Замечание: Если векторы рассматривать как какие-то характеристики объектов, то линейную зависимость можно интерпретировать как избыточность данных.

С помощью понятия линейной независимости вводится понятие **размерности** векторного пространства – это максимальное число линейно независимых векторов в нем.

Базис

В формальном определении вектора не присутствуют никакие его количественные интерпретации, а лишь две операции над векторами и 8 аксиом.

Откуда же появляются количественные измерения? Для того, чтобы это стало понятно, нужно ввести понятия **базиса**.

Базис – это конечный набор векторов в векторном пространстве, такой, что любой вектор этого пространства может быть единственным образом представлен в виде линейной комбинации векторов этого набора.

Вспомним один из примеров выше, где мы представляли прямоугольную систему координат на плоскости и единичные векторы (орты) $\mathbf{e}_1, \mathbf{e}_2$. $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ В данном примере мы разложили произвольный вектор \mathbf{a} следующим образом:

$$\begin{aligned} \mathbf{a} &= a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 \\ \mathbf{a} &= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_2 \end{pmatrix} \end{aligned}$$

$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$

Получается, что орты \mathbf{e}_1 , \mathbf{e}_2 являются базисом двумерного векторного пространства,

и с помощью линейной комбинации этих векторов мы можем единственным образом представить любой вектор этого пространства.

Возникает вопрос, данный базис является единственным в двумерном пространстве, или нет?

Ответ на этот вопрос – нет. На самом деле можно взять два любых вектора (**почти любых**), и они тоже будут являться базисом, при условии, что любой вектор можно разложить с помощью их линейной комбинации.

Пример.

Пусть у нас есть вектор $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ в базисе единичных орт, и мы хотим разложить его по другому базису $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ -3 \end{pmatrix}$:

$$\begin{aligned} & \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \\ & \begin{pmatrix} 2 \\ 3 \end{pmatrix} \end{aligned}$$

Откуда мы можем найти, что $\alpha_1 = -1$, $\alpha_2 = -1$:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ -3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Таким образом, вектор $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ в базисе единичных орт представляется как $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$ в базисе $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ -3 \end{pmatrix}$.

Но для базиса, как было сказано ранее, подойдет **не любой набор векторов**.

Например через базис $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -3 \\ 0 \end{pmatrix}$ нельзя разложить вектор $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$, поэтому данный набор векторов не является базисом.

В чем же принципиальная разница между этими базисами, и может ли базис двумерного пространства состоять, например, из большего или меньшего числа векторов, чем 2?

У линейной алгебры есть на это ответ: **Любые n линейно независимых векторов n -мерного векторного пространства образуют базис этого пространства.**

Именно из-за линейной зависимости векторов $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -3 \\ 0 \end{pmatrix}$ они не могут являться базисом двумерного пространства.

Что мы узнали?

- Определение вектора
- Операции с векторами
- Норма (длина) вектора
- Скалярное произведение
- Линейная независимость
- Базис

Ско бки

Дисклеймер: нам нужен инструментарий для работы с более абстрактными пространствами!

Символ Кронекера и остальные специальные тензоры – это удобные примитивы для низкоуровневой работы с матрицами и векторами. Физики смогли придумать поверх примитивов линейной алгебры отличный синтаксический сахар, т.н. нотацию Дирака. Но давайте начнем с определения гильбертова пространства.

Гильбертово пространство

Гильбертово пространство – это полное пространство, определяемое:

1. скалярным произведением $\langle u, v \rangle$, в простейшем частном случае – $\langle u, v \rangle = u^\dagger v$;
2. зафиксированной нормой вида $\|v\| = \sqrt{\langle v, v \rangle}$;
3. метрикой $d(u, v) = \|u - v\| = \sqrt{\langle u - v, u - v \rangle}$.

Полное пространство

При нахождении предела некоторой последовательности полное метрическое пространство дает возможность не думать о принадлежности этого предела данному пространству, то есть по своей сути разрешает предельный переход и позволяет доказать множество красивых теорем. Подробнее строгое определение можно посмотреть [тут](#).

Нотация Дирака, или Bra-Ket нотация

Напомним, что в гильбертовом пространстве для пространства $\langle H \rangle$ определено сопряженное пространство $\langle H^{\dagger} \rangle$. Тогда можно рассмотреть обозначение Ket для элемента $\langle v \rangle$ из $\langle H \rangle$ как вектор-столбец:

$\begin{aligned} \ket{v} = & \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} \end{aligned}$

и обозначение Bra для элемента $\langle u \rangle$ из сопряженного пространства $\langle H^{\dagger} \rangle$ как сопряженную вектор-строку

$\begin{aligned} \bra{u} = & \begin{bmatrix} u_1^* & u_2^* & \cdots & u_N^* \end{bmatrix} \end{aligned}$

Bra-ket $\langle u \rangle \langle v \rangle$ просто задает скалярное произведение между этими элементами:

$\begin{aligned} \begin{aligned} \bra{u} \ket{v} = & u_1^* v_1 + u_2^* v_2 + \cdots + u_N^* v_N = \begin{bmatrix} u_1^* & u_2^* & \cdots & u_N^* \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} \end{aligned} \end{aligned}$

Часто, например, требуется показать два граничных состояния кубита $\langle \ket{0} \rangle, \langle \ket{1} \rangle$, тогда мы можем записать их просто как вектор-столбцы: $\langle \ket{0} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \langle \ket{1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. В следующей лекции этот вопрос будет рассмотрен детальнее.

Внешнее произведение (outer-product)

Также имеет смысл и переставленная запись Ket-Bra, которая называется [внешним произведением](#):

$\begin{aligned} \begin{aligned} \ket{u} \bra{v} = & \begin{bmatrix} u_1^* & u_2^* & \cdots & u_N^* \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \begin{bmatrix} u_1^* v_1 + u_2^* v_2 + \cdots + u_N^* v_N \end{bmatrix} \end{aligned} \end{aligned}$

По сути это матрица $(N \times N)$, то есть новый оператор в гильбертовом пространстве. Не все перестановки имеют смысл, например, нельзя записать $\langle \bra{v} \bra{v} \rangle$ или $\langle \ket{u} \ket{v} \rangle$.

Мы вернемся к Ket-Bra чуть ниже, когда будем говорить про операторы проекции.

Эрмитов оператор

Оператор $\langle U \rangle$ называется [эрмитовым](#), если он удовлетворяет равенству $\langle (Uv, u) = (v, Uu) \rangle$ для всех $\langle u \rangle, \langle v \rangle$ из $\langle H \rangle$ или в матричном виде:

$\langle U = U^{\dagger} \rangle$

Унитарный оператор

[Унитарный оператор](#) $\langle \hat{U}: H \rightarrow H \rangle$ на гильбертовом пространстве $\langle H \rangle$ – это линейный оператор, который удовлетворяет следующему равенству $\langle (Uv, u) = (v, Uu) \rangle$ для любых $\langle v \in H \rangle, \langle u \in H \rangle$. Или в матричной форме:

$\langle \hat{U}^{\dagger} \hat{U} = \hat{U} \hat{U}^{\dagger} = I \rangle$

Напомним, что операция $\langle \hat{U}^{\dagger} \rangle$ (другое частое обозначение в работах – звездочка или $\langle H | \hat{U}^{\dagger} = \hat{U}^{\dagger} | H \rangle$) в матричных терминах является последовательным применением операции транспонирования и последующего комплексного сопряжения элементов этой матрицы $\langle \hat{U}^{\dagger} = \overline{\hat{U}^T}$ (порядок этих операций, естественно, не влияет на результат).

```
import numpy as np
import scipy as sp
from scipy import linalg

U_hat = np.array([
    [1 + 0j, 0 + 0j],
    [0 + 0j, 1 + 0j]
])

U_hat_star_byhands = U_hat.conj().T
U_hat_star_long = np.conjugate(np.transpose(U_hat))

print(np.allclose(U_hat_star_long, U_hat_star_byhands))
```

True

Также в Python для многих операций есть соответствующие методы вместо функций и их сокращения, например `U.transpose()` – то же самое, что `U.T`, а `U.conjugate()` – `U.conj()`.

Важное свойство, что любой эрмитов оператор \hat{U} можно привести к унитарному оператору с помощью взятия матричной экспоненты от матрицы оператора, умноженного на единицу:

$$[\hat{U} = e^{i\hat{U}}]$$

Давайте докажем быстро этот факт: для эрмитовых квадратных матриц можно определить матричные функции через спектральное разложение матрицы, которое в силу свойств эрмитовых матриц имеет вид: $U = S \Lambda S^{-1}$, где (S) – унитарная матрица перехода к тому базису, где изначальный оператор выражается диагональной матрицей. Тогда матричная функция $f(U) = S f(\Lambda) S^{-1}$, т.е. мы диагонализируем матрицу и применяем функцию ко всем диагональным элементам, а потом возвращаемся в исходный базис унитарным преобразованием. Соответственно, для экспоненты:

$$[e^U = S e^{\Lambda} S^{-1}]$$

Note

В качестве упражнения для самопроверки можете показать, что такое определение эквивалентно определению через степенной ряд: $(e^U = \sum_{k=0}^{\infty} \frac{1}{k!} U^k)$

Доказательство: Проверим определение унитарной матрицы: $(\hat{U}^\dagger \hat{U}) = (S \exp(i\Lambda) S^\dagger)$ $(S^\dagger \exp(i\Lambda) S^\dagger)$. Пользуясь тем, что эрмитово сопряжение переставляет матрицы и является обратной операцией к себе же, а также тем, что (S) – унитарная матрица (то есть $(S \cdot S^\dagger = I)$), получаем:

$$[(S \exp(i\Lambda) S^\dagger)^\dagger S \exp(i\Lambda) S^\dagger = \exp(-i\Lambda) S^\dagger S \exp(i\Lambda) S^\dagger]$$

Заметим, что теперь у нас уже скалярная экспонента, которая применяется к элементам диагональной матрицы, и мы можем воспользоваться тем, что произведение экспонент превращается в экспоненту от суммы степеней:

$$[S \exp(-i\Lambda) + i\Lambda S^\dagger S \exp(i\Lambda) S^\dagger = S \exp(O) S^\dagger = S I S^\dagger = I]$$

В конце мы еще раз воспользовались тем, что (S) – унитарная. Абсолютно так же доказывается, что $(\hat{U} \hat{U}^\dagger = I)$.

Note

Кстати, любая матрица вида $(H^\dagger H)$ является эрмитовой.

Давайте продемонстрируем доказанный факт на примере матрицы дискретного преобразования Фурье без нормировочного коэффициента $(\frac{1}{N})$, $(N=3)$, преобразованной к (DD^\dagger) :

```
N = 3
w = np.sqrt(np.exp(-1j * 2 * np.pi / N))

D = np.array([
    [1, 1, 1],
    [1, w, w ** 2],
    [1, w ** 2, w ** 4]
])

U = D @ D.conj().T
print(f"U = {U}")

print(f"D = {D}")
print(f"np.allclose(U @ U.T.conj(), np.eye(N)) = {np.allclose(U @ U.T.conj(), np.eye(N))} # no")
print(f"\nnp.allclose(U_hat @ U_hat.conj().T, np.eye(N)) = {np.allclose(U_hat @ U_hat.conj().T, np.eye(N))}")
print(f"\nnp.allclose(U_hat.conj().T @ U_hat, np.eye(N)) = {np.allclose(U_hat.conj().T @ U_hat, np.eye(N))}")
```

```

U = array([[ 3.0000000e+00+0.0000000e+00j,  1.0000000e+00+1.73205081e+00j,
           -1.11022302e-16+7.77156117e-16j],
           [ 1.0000000e+00-1.73205081e+00j,  3.0000000e+00+0.0000000e+00j,
           1.0000000e+00+1.73205081e+00j],
           [-1.11022302e-16-7.77156117e-16j,  1.0000000e+00-1.73205081e+00j,
           3.0000000e+00+0.0000000e+00j]])
D = array([[ 1. +0.j          ,  1. +0.j          ,  1. +0.j          ],
           [ 1. +0.j          ,  0.5-0.8660254j, -0.5-0.8660254j],
           [ 1. +0.j          , -0.5-0.8660254j, -0.5+0.8660254j]])
np.allclose(U @ U.T.conj(), np.eye(N)) = False
np.allclose(U_hat @ U_hat.conj().T, np.eye(N)) = True
np.allclose(U_hat.conj().T @ U_hat, np.eye(N)) = True

```

Пример: оператор-проектор

Оператором проекции является оператор $\langle P \rangle$ со свойством $\langle P^2 = P \rangle$.

Покажем, что Ket-Bra вида $\langle \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle \rangle$ обладает этим свойством.

$$\langle \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle \rangle^2 = \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle = \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle = \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle$$

поскольку вектор состояния – нормированный: $\langle \langle \text{bra}\{\Psi\} | \text{ket}\{\Psi\} \rangle \rangle = 1$.

Оператор Ket-Bra с вектором состояния $\langle \langle \text{ket}\{\Psi\} \rangle \rangle$, то есть $\langle \langle \text{ket}\{\Psi\} | \text{bra}\{\Psi\} \rangle \rangle$ рассматривается во [вводной лекции про кубиты](#), а также пригодится позже, когда речь зайдет о [смешанных состояниях](#).

Пример: оператор поворота

Оператором поворота по оси вращения $\langle v=(x,y,z) \rangle$ на угол $\langle \theta \rangle$ является

$$\begin{aligned} M(v, \theta) = \begin{pmatrix} \cos \theta & (1 - \cos \theta)x & (1 - \cos \theta)y \\ (1 - \cos \theta)x & \cos \theta & (-\sin \theta)z \\ (1 - \cos \theta)y & (\sin \theta)z & \cos \theta \end{pmatrix} = \\ \begin{pmatrix} \cos \theta & (1 - \cos \theta)x & (1 - \cos \theta)y \\ (1 - \cos \theta)x & \cos \theta & (-\sin \theta)z \\ (1 - \cos \theta)y & (\sin \theta)z & \cos \theta \end{pmatrix} \end{aligned}$$

Например, матрица поворота относительно оси $\langle x \rangle$ на $\langle 90^\circ \rangle$: $\langle (x=1), (y=0), (z=0), (\theta = \frac{\pi}{2}) \rangle$, будет иметь вид:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -\cos(\frac{\pi}{2}) & -\sin(\frac{\pi}{2}) \\ 0 & \sin(\frac{\pi}{2}) & 0 & \cos(\frac{\pi}{2}) \end{pmatrix}$$

Если у нас больше измерений, то по тем измерениям, которые не затрагиваются этим поворотом, у нас в строке и столбце стоят нули, кроме пересечения строки и столбца, отвечающих за это измерение - там стоит единица.

Операторы поворота очень важны в квантовых вычислениях. Они рассматриваются отдельно в [лекции про квантовые гейты](#) и далее используются в [вариационных квантовых схемах](#) для кодирования классических данных в квантовые операторы.

Пример: оператор дифференцирования

В пространстве многочленов $\langle P \rangle$ с базисом $\langle \langle 1, t, t^2, \dots, t^n \rangle \rangle$ можно задать оператор дифференцирования $\langle \mathcal{D} \rangle$ в виде матрицы:

$$\begin{aligned} D = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix} \end{aligned}$$

Тогда производная многочлена $\langle p = a_0 + a_1 t + \dots + a_n t^n \rangle = \begin{pmatrix} a_0 & a_1 & \dots & a_n \end{pmatrix}$

$$\begin{aligned} \langle \mathcal{D}(p) \rangle = \begin{pmatrix} a_0 & a_1 & \dots & a_n \end{pmatrix} D \begin{pmatrix} 1 & t & \dots & t^n \end{pmatrix} = \\ a_1 + 2a_2 t + \dots + n a_n t^{n-1} \end{aligned}$$

❶ Про разницу между оператором и матрицей преобразования

Мы с вами рассматривали операторы через конечные матрицы, но на самом деле гильбертово пространство было придумано как раз, чтобы можно было работать с бесконечномерными векторами и применять непрерывные операторы. Можно считать, что есть некоторый дуализм между непрерывным оператором и пределом бесконечной матрицы. Иногда удобнее работать с матрицей, а иногда – с абстрактным оператором.

Произведение Кронекера

Давайте рассмотрим еще одну интересную операцию, которая называется матричным тензорным произведением (является тензорным произведением для линейных операторов) или произведением Кронекера.

Проще всего его необходимость можно продемонстрировать на примере двух игр: Орел/Решка и бросок кубика. Мы можем записать состояния этих игр через вероятности событий и давайте возьмем монетку со смешанным центром тяжести и такой же кубик:

$\text{\text{coin}} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$ для нашей монетки и $\text{\text{dice}} = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$ для нашей игральной кости. Тогда если мы захотим сыграть в игру, когда сначала подкидывается монетка, а потом - игральный кубик, нам будет удобно записать это в виде либо очень длинного вектора:

$$\begin{aligned} \text{\text{game}}_{\text{\text{vec}}} = & \begin{bmatrix} \frac{1}{3} \times \frac{1}{6} & \frac{1}{3} \times \frac{1}{6} \\ \frac{2}{3} \times \frac{1}{6} & \frac{2}{3} \times \frac{1}{6} \end{bmatrix} \\ & \times \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \times \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \end{aligned}$$

Либо в виде матрицы, где по строкам будут события монетки, а по столбцам - кубика:

$$\begin{aligned} \begin{bmatrix} \text{\text{game}}_{\text{\text{matrix}}} \end{bmatrix} = & \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \times \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \\ & \times \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{aligned}$$

С помощью произведения Кронекера (или, повторимся, – матричного тензорного произведения) похожие огромные вектора и матрицы можно очень компактно записать:

$$\text{\text{game}}_{\text{\text{vec}}} = \text{\text{coin}} \otimes \text{\text{dice}} \quad \text{\text{game}}_{\text{\text{matrix}}} = \text{\text{coin}}^T \otimes \text{\text{dice}}$$

В общем случае,

$$\begin{aligned} \begin{bmatrix} \text{\text{split}} \end{bmatrix} A \otimes B = & \begin{bmatrix} a_{11} & \cdots & a_{1n} & \vdots & \ddots & \vdots & a_{m1} & \cdots & a_{mn} \end{bmatrix} \times \begin{bmatrix} b_{11} & \cdots & b_{12} & \cdots & b_{1n} & \vdots & b_{21} & \cdots & b_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} & \vdots & \ddots & \vdots & b_{(m-1)1} & \cdots & b_{(m-1)n} \end{bmatrix} \\ = & \begin{bmatrix} a_{11}b_{11} & \cdots & a_{11}b_{12} & \cdots & a_{11}b_{1n} & \vdots & a_{11}b_{21} & \cdots & a_{11}b_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & \cdots & a_{mn}b_{mn} & \vdots & \ddots & \vdots & b_{(m-1)1}a_{11} & \cdots & b_{(m-1)n}a_{11} \end{bmatrix} \end{aligned}$$

Основные его свойства вы можете прочитать в статье: [Произведение Кронекера](#)

Есть и другие нужные тензорные операции, например, чуть больший список вы можете найти в [этой статье](#) или в рекомендованной литературе по квантовой механике.

Рекомендованная литература

- Quantum Mechanics: The Theoretical Minimum Illustrated Edition by Leonard Susskind, Art Friedman [[SF14](#)] или его русский перевод [[15b](#)]. Книга сочетает в себе довольно строгий формализм вместе с интуицией квантовой механики. Главы построены в виде конкретных примеров или задач, которые понимают, зачем та или иная теория необходима.
- Mathematics for machine learning by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong [[DFO20](#)] - отличная книжка чтобы освежить воспоминания о базовых алгоритмах в машинном обучении и о математике, которая применяется в них. [Доступна бесплатная электронная версия](#).
- Deep Learning by Ian Goodfellow, Yoshua Bengio, Aaron Courville [[GBC16](#)] - отличная книжка, чтобы разобраться в основах глубокого обучения, чем-то может заменить предыдущую книгу, отличный старт, если вы хотите разобраться в автоэнкодерах или в других нейросетевых моделях. [Доступна бесплатная электронная версия](#).

Что мы узнали

- произвол со скобками
- Гильбертовы пространства

- Эрмитовий оператор
- Унітарний оператор
- Примеры различных операторов

Матрицы

Определение

Матрица – это математический объект, представляющий собой прямоугольную таблицу чисел.

У каждой матрицы есть размер, который характеризуется двумя числами – количеством строк и количеством столбцов. Матрицу, состоящую из (n) строк (m) столбцов будем называть матрицей размера $(n \times m)$. Пространство вещественных матриц $(n \times m)$ обозначается $(\mathbf{R}^{n \times m})$. Мы будем рассматривать матрицу либо как набор (n) вектор-строк, либо как набор (m) вектор-столбцов.

Замечание: Вектор-строку из (n) элементов можно рассматривать как матрицу размера $(1 \times n)$, а аналогичного размера вектор-столбец – как матрицу размера $(n \times 1)$.

Рассмотрим произвольную матрицу (\mathbf{A}) :

$$\begin{aligned} \begin{aligned} \mathbf{A} = & \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \end{aligned} \end{aligned}$$

Элемент матрицы (\mathbf{A}) , стоящий на пересечение строки (i) и столбца (j) будем обозначать (a_{ij}) .

В машинном обучении часто приходится иметь дело с матрицей объекты-признаки. Объекты, как и признаки – это набор числовых характеристик (точки в многомерном пространстве), которые обычно рассматриваются как векторы.

Операции с матрицами

Аналогично векторным операциям, для матриц определены операции **сложения и умножения на число**:

- сложение:

$$\begin{aligned} \begin{aligned} \mathbf{A} + \mathbf{B} = & \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2m} + b_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \dots & a_{nm} + b_{nm} \end{pmatrix} \end{aligned} \end{aligned}$$

- умножение матрицы на число (скаляр):

$$\begin{aligned} \begin{aligned} \lambda \cdot \mathbf{A} = & \lambda \cdot \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} = \begin{pmatrix} \lambda \cdot a_{11} & \lambda \cdot a_{12} & \dots & \lambda \cdot a_{1m} \\ \lambda \cdot a_{21} & \lambda \cdot a_{22} & \dots & \lambda \cdot a_{2m} \\ \dots & \dots & \dots & \dots \\ \lambda \cdot a_{n1} & \lambda \cdot a_{n2} & \dots & \lambda \cdot a_{nm} \end{pmatrix} \end{aligned} \end{aligned}$$

- транспонирование (строки и столбцы меняются местами):

$$\begin{aligned} \begin{aligned} \mathbf{A}^T = & \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \xrightarrow{\text{матрица } A^T} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix} \end{aligned} \end{aligned}$$

Но наибольший интерес представляет операция умножения матриц, которая определяется не самым интуитивным способом.

Рассмотрим операцию скалярного произведения как умножение матрицы размера $(1 \times n)$ (вектор-строка) на матрицу размера $(n \times 1)$ (вектор-столбец). Результатом данной операции получается матрица размера (1×1) (скаляр).

Таким образом для получения значения (c_{11}) результирующей матрицы (\mathbf{C}) мы скалярно умножили 1-ю строку $(\mathbf{a}_{1:})$ матрицы (\mathbf{A}) (вектор) на 1-й столбец $(\mathbf{b}_{:1})$ матрицы (\mathbf{B}) (вектор).

Из примере выше мы можем прийти к следующим заключениям:

- для умножения матриц необходимо, чтобы количество столбцов первой матрицы было равно количеству строк второй, иначе нельзя посчитать скалярное произведение

- операция умножения матриц не является коммутативной, то есть $\mathbf{A} \mathbf{B} \neq \mathbf{B} \mathbf{A}$

Развивая логику примера выше мы можем перейти к определению операции умножения матриц. Рассмотрим матрицы $\mathbf{A} \in \mathbf{R}^{n \times p}$ и $\mathbf{B} \in \mathbf{R}^{p \times m}$. Результатом произведения этих матриц будет матрица $\mathbf{C} = \mathbf{A} \mathbf{B} \in \mathbf{R}^{n \times m}$, где значение c_{ij} получится путем скалярного произведения i -й строки $\mathbf{a}_{i:}$ матрицы \mathbf{A} (вектор) на j -й столбец $\mathbf{b}_{:j}$ матрицы \mathbf{B} (вектор):

$$c_{ij} = \mathbf{a}_{i:} \cdot \mathbf{b}_{:j} = a_{i1} b_{1j} + a_{i2} b_{2j} + \dots + a_{ip} b_{pj}$$

Замечание 1: В результате умножения квадратных матриц размера $(n \times n)$ получается квадратная матрица того же размера.

Замечание 2: Частными случаями матричного умножения являются операции умножения вектор-строки на матрицу и матрицы на вектор-столбец, в результате которых получаются вектор-строка и вектор-столбец соответственно.

Рассмотрим пример, почему удобно представлять операцию умножения матриц именно таким образом. Пусть имеется система линейных уравнений:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m = b_2 \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nm}x_m = b_n \end{cases}$$

Представим данную систему в матричном виде:

$$\begin{aligned} \mathbf{A} \mathbf{x} = \mathbf{b} \Leftrightarrow & \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \end{aligned}$$

Норма матрицы

Понятие матричной нормы несколько сложнее, чем векторной. Обычно принято рассматривать так называемые **операторные нормы**. Такая норма показывает насколько максимально растягивается произвольный вектор \mathbf{x} при отображении $y = \mathbf{Ax}$. Формально операторная норма определяется следующим образом:

$$\| \mathbf{A} \|_p = \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p}$$

Существует множество матричных норм, но наиболее полезной является **спектральная норма**. Опуская математические выкладки, можно показать, что спектральная норма равна максимальному собственному числу матрицы $\mathbf{A}^T \mathbf{A}$:

$$\| \mathbf{A} \|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})}$$

Про собственные числа поговорим позже, а пока вернемся к нормам. Иногда также рассматривают **поэлементные нормы** матриц. Такая норма определяется тем же образом, что и для векторов. Наиболее распространенной из данного вида является **Евклидова норма**, она же **норма Фробениуса**:

$$\| \mathbf{A} \|_F = \sqrt{\sum_{i,j} a_{ij}^2}$$

Ранг матрицы

Вспомним, что матрицу можно рассматривать либо как набор вектор-строк, либо как набор вектор-столбцов, а для набор векторов определено понятие линейной независимости.

Таким образом мы приходим двум следующим понятиям:

- **строчный ранг** матрицы – это максимальное число линейно независимых строк (вектор-строк)
- **столбцовый ранг** матрицы – это максимальное число линейно независимых столбцов (вектор-столбцов)

Тут мы можем задаться вопрос о том, равны ли эти ранги между собой, и если равны, то при каких условиях, и на помощь нам приходит **фундаментальная теорема «о ранге матрицы»**, которая говорит, что ранг, посчитанный по строкам всегда равен рангу, посчитанному по столбцам. Поэтому обычно говорят просто о ранге матрицы.

Линейные преобразования

В рамках школьной программы все знакомы с понятием функции. Функция – это соответствием между элементами двух множеств.

Например, линейная функция вида $y = f(x) = kx + b$ отображает элементы множества вещественных чисел $x \in \mathbf{R}$ в себя $y \in \mathbf{R}$:

$$f: \mathbf{R} \rightarrow \mathbf{R}$$

Парабола $y = f(x) = x^2$ отображает элементы множества вещественных чисел $\{x \in \mathbf{R}\}$ во множество неотрицательных вещественных чисел $\{y \in \mathbf{R} \mid y \geq 0\}$:

$$f: \mathbf{R} \rightarrow \mathbf{R} \cup \{0\}$$

Среди всего множества функций есть определенный класс функций, называемых линейными. Этот класс является самым простым среди множества всех функций. Основным свойством линейных функций является пропорциональность $y(x)$: изменение значения функции y пропорционально изменению значения аргумента x . Графиком линейной функции является прямая.

Обобщая идею линейных функций на матрицы и вектора мы можем прийти к понятиям **линейных отображений** и **линейных преобразований**.

Линейное отображение – это обобщение линейной функции, которое принимает на вход вектор и возвращает вектор. Вектор может быть как той же размерности, так и другой.

Формально, линейное отображение – это отображение множества элементов одного векторного пространства V в другое W

$$f: V \rightarrow W,$$

удовлетворяющее так называемым [условиям линейности](#).

Если V и W — это одно и то же векторное пространство, то f называют не просто линейным отображением, а **линейным преобразованием**.

Отображения и преобразования удобно рассматривать как движение в векторном пространстве.

Ранее упоминалось, что при умножении вектора на матрицу мы получаем вектор, поэтому матрица задает преобразование, которое, если проверить условия линейности, является линейным. Соответственно, **матрица задает линейное отображение**.

Например, матрица $A \in \mathbf{R}^{n \times m}$ задает преобразование из векторного пространства размерности n в векторное пространство размерности m .

Замечание 1: Для того, чтобы матрица задавала линейное преобразование, она должна быть квадратной, т.к. в результате умножение вектора на квадратную матрицу его размерность не меняется. Таким образом вектор, как элемент векторного пространства, отображается в то же векторное пространство.

Замечание 2: Пусть имеются две матрицы $A \in \mathbf{R}^{n \times p}$ и $B \in \mathbf{R}^{p \times m}$. Умножив матрицу A на матрицу B мы получим матрицу $C \in \mathbf{R}^{n \times m}$. Матрица C задает линейное отображение, равное последовательному применению отображений, задаваемых матрицами A и B . Таким образом операция умножения матриц имеет простую интерпретацию - последовательное применение линейных отображений.

Рассмотрим некоторые примеры линейных преобразований на плоскости, задаваемых матрицами размера 2×2 :

- Вращение / поворот:
 - на 90 градусов против часовой стрелки: $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
 - на угол θ против часовой стрелки: $A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$
- Отражение:
 - относительно оси x : $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
 - относительно оси y : $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$
 - относительно линии, образующей угол θ с началом координат: $A = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$
- Масштабирование:
 - в 2 раза по всем направлениям: $A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$
- Проекция:
 - на ось y : $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$

Обратная матрица

Рассмотрим особую матрицу линейного преобразования, которая переводит каждый вектор сам в себя, т.е. никак его не изменяет. Данное преобразование задается так называемой **единичной матрицей**:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Произведение любой матрицы / вектора и единичной матрицы подходящего размера равно самой матрице / вектору (аналог единицы для операции умножения чисел):

$$\forall \mathbf{A} \mathbf{E} = \mathbf{E} \mathbf{A} = \mathbf{A}$$

Появление такого элемента в пространстве вещественных матриц наводит на мысль о том, что для каждой матрицы должен существовать обратный элемент:

$$\forall \mathbf{A} \mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{I}$$

Замечание: Аналогично операции умножения чисел, если $(a \cdot 1 = 1 \cdot a = a)$, то существует (a^{-1}) , такой, что $(a \cdot a^{-1} = a^{-1} \cdot a = 1)$.

Теория линейной алгебры говорит о том, что обратная матрица существует только для **невырожденных квадратных матриц**, строки и столбцы которых линейно независимы.

Те линейные преобразования, которые были приведены в примерах выше задаются невырожденными квадратными матрицами, поэтому для них существуют обратные преобразования, которые тоже будут линейными.

Например, обратными к преобразованиям вращения / поворота, отражения и масштабирования являются эти же преобразования: вращение / поворот, отражение и масштабирование соответственно.

Замечание: Существуют разные алгоритмы нахождения обратной матрицы, которые не будут нами рассматриваться, но с ними можно ознакомиться [здесь](#).

Будем называть линейное преобразование **вырожденным**, если оно задается вырожденной матрицей.

Соответственно, возникает вопрос, чем же так особенны вырожденные линейные преобразования, если для них не существует обратного преобразования?

Невырожденное линейное преобразование устанавливает взаимно-однозначное соответствие между входными и выходными векторами. Вырожденное преобразование этим свойством не обладает.

Это связано с тем, вырожденные линейные преобразования переводят некоторые ненулевые вектора в нулевые (для невырожденных линейных преобразований только нулевой вектор переходит в нулевой):

$$\begin{aligned} \mathbf{A} \mathbf{x} = 0, \mathbf{x} \neq 0 &= \begin{pmatrix} 0 & 0 & \dots & 0 \end{pmatrix} \end{aligned}$$

В связи с этим одному выходному вектору (нулевому) соответствует несколько входных векторов, поэтому не существует обратного преобразования.

Определитель

Введем еще одно из ключевых понятий линейной алгебры – **определитель**.

Определитель – это скалярный показатель (число) **квадратной матрицы** $\left(f: \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R} \right)$, который позволяет охарактеризовать некоторые свойства этой матрицы и линейного преобразования, заданного этой матрицей.

Определитель матрицы \mathbf{x} обозначается как: $\det(\mathbf{x}) = |\mathbf{x}| = \Delta(\mathbf{x})$

Формально понятие определителя может быть введено несколькими способами:

- через перестановки
- через свойства определителя
- геометрически

Замечание: В данной лекции не будет рассказано, как посчитать определитель. Ознакомиться с алгоритмом расчета и формальным выводом определителя можно по [ссылке](#).

Определитель обладает следующими важными свойствами – он **равен нулю** тогда и только тогда, когда **строки (столбцы) матрицы линейно зависимы**.

Из этого можно сделать следующие выводы:

- матрица вырожденная тогда и только тогда, когда ее определитель равен 0
- линейное преобразование вырожденное тогда и только тогда, когда определитель матрицы, задающей данное преобразование, равен 0
- если определитель матрицы равен 0, то для данной матрицы не существует обратной

Существуют эффективные алгоритмы вычисления определителя квадратной матрицы, которые позволяют проверять, является матрица вырожденной.

Собственные числа и векторы

Для некоторых линейных преобразований существуют такие векторы, которые после данного преобразования либо растягиваются, либо сжимаются (тот же вектор, умноженный на некоторое число). Такие векторы называются **собственными векторами** линейного преобразования, а числа, на которые данные векторы умножаются – **собственными значениями**.

Рассмотрим произвольную матрицу линейного преобразования \mathbf{A} . Формально понятия собственных векторов и собственных значений вводятся через следующее равенство: $\mathbf{A} \mathbf{x} = \lambda \mathbf{x}$

Векторы \mathbf{x} , удовлетворяющие данному равенству являются собственными векторами линейного преобразования \mathbf{A} , а соответствующие данным векторам значения λ – собственными числами.

Линейное преобразование может как не иметь собственных векторов (например, поворот в двумерном пространстве), или иметь n собственных векторов с различными собственными значениями.

Собственные числа и собственные значения играют большую роль в линейной алгебре, так как многие соотношения, связанные с линейными преобразованиями, существенно упрощаются в системе координат, построенной на базисе из собственных векторов линейного преобразованиями.

Множество собственных значений линейного преобразованиями (**спектр оператора**) характеризует важные свойства оператора без привязки к какой-либо конкретной системе координат.

Также собственные векторы используются в Методе Главных Компонент, который предназначен для уменьшения размерности данных с потерей наименьшего количества информации.

Что мы узнали?

- Определение матрицы
- Операции с матрицами
- Норма матрицы
- Ранг матрицы
- Линейные преобразования
- Обратная матрица
- Определитель матрицы
- Собственные числа и векторы матрицы

Numpy

Numpy – это широко используемая библиотека для вычислений с многомерными массивами. API большей частью вдохновлен **MATLAB** (великая и ужасная среда, язык и IDE для матричных вычислений), а теперь сам является примером для подражания API различных вычислительных пакетов. Более последовательный гайд стоит посмотреть на [сайте библиотеки](#).

Массивы

```
import numpy as np

a = np.array([1, 2, 3]) # создадим вектор
print(f"{a = }")

b = np.zeros((2, 2))
print(f"{b = }")

c = np.eye(3)
print(f"{c = }")

q = np.random.random((1, 20))
print(f"{q = }")
```

```

a = array([1, 2, 3])
b = array([[0., 0.],
           [0., 0.]])
c = array([[1., 0., 0.],
           [0., 1., 0.],
           [0., 0., 1.]])
q = array([[0.35064708, 0.1221341 , 0.9730359 , 0.11779796, 0.6993172 ,
           0.64868333, 0.15322228, 0.28524572, 0.20437489, 0.38330839,
           0.11098454, 0.33769526, 0.26929776, 0.55379138, 0.13999579,
           0.08886012, 0.25081757, 0.75233615, 0.49817054, 0.92670358]])
```

Арифметические операции

Для удобства использования `np.ndarray` определены арифметические операторы, так чтобы соответствовать ожиданиям:

```

a = np.array([1, 2, 3])
b = np.array([-1, 3, 4])

diff = a - b
print(f"{diff = }")

mult = a * b
print(f"{mult = }")

scalar_mult = a @ b
print(f"{scalar_mult = }")
```

```

diff = array([ 2, -1, -1])
mult = array([-1,  6, 12])
scalar_mult = 17
```

Indexing, slicing and sugar

`Numpy` поддерживает, кажется, все разумные варианты индексации:

```

a = np.arange(16).reshape(4, 4)
print(f"{a = }")

# просто по индексам
print(f"\n{a[0, 1] = }")
print(f"{a[0][1] = }")

# по слайсам
print(f"\n{a[0, 1:3] = }")
print(f"{a[2] = }")
print(f"{a[2, :] = }")
print(f"{a[2, ...] = }")

# по маске
mask = (a % 3 == 0)
print(f"\n{mask = }")
print(f"{a[mask] = }")

first_rows = np.array([True, True, False, False])
print(f"\n{a[first_rows] = }")
```

```

a = array([[ 0,  1,  2,  3],
          [ 4,  5,  6,  7],
          [ 8,  9, 10, 11],
          [12, 13, 14, 15]])

a[0, 1] = 1
a[0][1] = 1

a[0, 1:3] = array([1, 2])
a[2] = array([ 8,  9, 10, 11])
a[2, :] = array([ 8,  9, 10, 11])
a[2, ...] = array([ 8,  9, 10, 11])

mask = array([[ True, False, False,  True],
              [False, False,  True, False],
              [False,  True, False, False],
              [ True, False, False,  True]])
a[mask] = array([ 0,  3,  6,  9, 12, 15])

a[first_rows] = array([[0, 1, 2, 3],
                      [4, 5, 6, 7]])

```

Для работы с размерностями часто используются еще три конструкции: `None`, `...` (ellipsis, многоточие) и `:` (двоеточие).

```

a = np.arange(16).reshape(4, 4)
print(f"{a = }")

# None добавляет ось размерности 1
print(f"\n{a[None].shape = }")
print(f"{a[:, :, None].shape = }")

# : превращается в slice (None), берет все элементы вдоль размерности
print(f"\n{a[2, :] = }")
print(f"\n{a[2, 0:None] = }")

# ... ellipsis, превращается в необходимое число двоеточий :::::
print(f"\n{a[... ] = }")

# также ... удобен когда мы не знаем настоящий шейп массива или нужно не трогать
# несколько подряд идущих размерностей
z = np.arange(27).reshape(3, 3, 3)
print(f"\n{z[0, ..., 1] = }")
print(f"\n{z[0, :, 1] = }")

```

```

a = array([[ 0,  1,  2,  3],
          [ 4,  5,  6,  7],
          [ 8,  9, 10, 11],
          [12, 13, 14, 15]])

a[None].shape = (1, 4, 4)
a[:, :, None].shape = (4, 4, 1)

a[2, :] = array([ 8,  9, 10, 11])
a[2, 0:None] = array([ 8,  9, 10, 11])

a[...] = array([[ 0,  1,  2,  3],
               [ 4,  5,  6,  7],
               [ 8,  9, 10, 11],
               [12, 13, 14, 15]])

z[0, ..., 1] = array([1, 4, 7])
z[0, :, 1] = array([1, 4, 7])

```

В целом, в NumPy очень здорово реализованы методы `__getitem__`/`__setitem__`.

```

a = np.array([1, 2, 3])
element = a[2]
print(f"{element = }")

a[2] = 5
print(f"\n{a = }")

```

```

element = 3
a = array([1, 2, 5])

```

Кроме того, мы можем делать индексацию по заданному условию с помощью `np.where`

```
# создадим вектор
a = np.array([2, 4, 6, 8])

selection = np.where(a < 5)
print(f"{selection = }")

# дополнительно мы можем передать два значения или вектора, при выполнении условия
# выбираются элементы из первого значения/вектора, при невыполнении - из второго
a2 = np.where(a < 5, 2, a * 2)
print(f"{a2 = }")

# np.where работает и с многомерными массивами
b = np.array([[8, 8, 2, 6], [0, 5, 3, 4]])
b_mult = np.where(b < 4, b, 1)
print(f"{b_mult = }")
```

```
selection = (array([0, 1]),
a2 = array([ 2,  2, 12, 16])
b_mult = array([[1, 1, 2, 1],
[0, 1, 3, 1]])
```

Broadcasting

Что происходит, если мы хотим производить арифметические операции с массивами разных размеров?

```
a = np.array([1, 2, 3])
k = 2
broad = a * k
print(f"{broad = }")
```

```
broad = array([2, 4, 6])
```

С точки зрения математики, ничего интересного тут не происходит: мы подразумевали умножение всего вектора на скаляр. Однако матричные операции в numpy справляются и с менее очевидными случаями, например при сложении или вычитании вектора и скаляра:

```
a = np.array([1, 2, 3])
k = 2
broad = a - k
print(broad)
```

```
[-1  0  1]
```

В numpy приняты следующие правила работы с массивами разного размера:

1. Размерности сравниваются справа налево
2. Два массива совместимы в размерности, если она одинаковая, либо у одного из массивов единичная.
3. Вдоль отсутствующих размерностей происходит расширение повторением (`np.repeat`).



⚠️ Attention

Be aware Автоматический броадкастинг легко приводит к ошибкам, так что лучше делать его самостоятельно в явной форме.

Операции с плавающей точкой

Отдельно стоит поговорить про числа с плавающей точкой. Число с плавающей точкой (или число с плавающей запятой) — экспоненциальная форма представления вещественных (действительных) чисел, в которой число хранится в виде мантиссы и порядка (показателя степени). При этом число с плавающей точкой имеет фиксированную относительную точность и изменяющуюся абсолютную. В результате одно и то же значение может выглядеть по-разному, если хранить его с разной точностью.

```
f16 = np.float16("0.1")
f32 = np.float32(f16)
f64 = np.float64(f32)
print(f"{f16 = }, {f32 = }, {f64 = }")
print(f"{f16 == f32 == f64 = }")

f16 = np.float16("0.1")
f32 = np.float32("0.1")
f64 = np.float64("0.1")
print(f"{f16 = }, {f32 = }, {f64 = }")
print(f"{f16 == f32 == f64 = }")
```

```
f16 = 0.1, f32 = 0.099975586, f64 = 0.0999755859375
f16 == f32 == f64 = True
f16 = 0.1, f32 = 0.1, f64 = 0.1
f16 == f32 == f64 = False
```

Из-за этого для сравнения массивов с типом float используют `np.allclose`.

```
print(f"{np.allclose([1e10,1e-7], [1.00001e10,1e-8]) = }")
print(f"{np.allclose([1e10,1e-8], [1.00001e10,1e-9]) = }")
```

```
np.allclose([1e10,1e-7], [1.00001e10,1e-8]) = False
np.allclose([1e10,1e-8], [1.00001e10,1e-9]) = True
```

NumPy и линейная алгебра

В **Numpy** много удобных функций, которые позволяют упростить код. Приведем несколько примеров:

```
# матрица с единицами по диагонали и с нулями в остальных ячейках.  
print(f"np.eye(2, dtype=int) = {np.eye(2, dtype=int)}")  
# есть возможность указать индекс диагонали  
print(f"np.eye(3, k=-1, dtype=int) = {np.eye(3, k=-1, dtype=int)}")  
  
# в Numpy есть свой генератор случайных чисел и векторов  
print(f"np.random.beta(1, 2) = {np.random.beta(1, 2)}")  
print(f"np.random.randint(1, 5, (2, 3)) = {np.random.randint(1, 5, (2, 3))}")  
  
# Numpy позволяет заменить значений основной диагонали матрицы.  
# Внимание, эта функция работает in-place  
a = np.random.randint(1, 5, (3, 3))  
print(f"a = {a}")  
np.fill_diagonal(a, 4)  
print(f"a = {a}")  
  
# Можно сделать и наоборот - получить вектор значений диагонали матрица  
print(f"np.diag(a) = {np.diag(a)}")
```

```
np.eye(2, dtype=int) = array([[1, 0],  
[0, 1]])  
np.eye(3, k=-1, dtype=int) = array([[0, 0, 0],  
[1, 0, 0],  
[0, 1, 0]])  
np.random.beta(1, 2) = 0.06744418401494583  
np.random.randint(1, 5, (2, 3)) = array([[2, 4, 2],  
[1, 4, 4]])  
a = array([[4, 3, 1],  
[1, 2, 3],  
[4, 1, 2]])  
a = array([[4, 3, 1],  
[1, 4, 3],  
[4, 1, 4]])  
np.diag(a) = array([4, 4, 4])
```

Решение систем линейных уравнений

Numpy позволяет решить систему линейных уравнений.

```
a = np.array([[7, 4], [9, 8]])  
b = np.array([5, 3])  
solution = np.linalg.solve(a, b)  
print(solution)
```

```
[ 1.4 -1.2]
```

Обращение матриц

Numpy дает возможность выполнить операцию обращения матриц.

```
a = np.array([[1., 2.], [3., 4.]])  
inv = np.linalg.inv(a)  
print(inv)
```

```
[[ -2.   1. ]  
 [ 1.5 -0.5]]
```

Собственные векторы и числа

Вычисление собственных векторов и чисел.

```
print(np.linalg.eig(np.diag((1, 2, 3))))
```

```
(array([1., 2., 3.]), array([[1., 0., 0.],  
[0., 1., 0.],  
[0., 0., 1.]]))
```

Мы вкратце рассмотрели

- основы работы с [NumPy](#)
- индексацию в массивах
- broadcasting массивов [NumPy](#)
- операции с плавающей точкой
- [NumPy](#) и примитивы линейной алгебры

О блоке “Машинное обучение”

Этот блок рассказывает о том, что общего у разных задач машинного обучения и как из основных компонентов, подобно паззлу, складываются различные применения моделей машинного обучения. Основная наша задача – чтобы у читателя даже не знакомого с машинным обучением возникло понимание того, как это все работает и как “сложить паззл” в новой задаче, будь то кластеризация новостей, детекция лиц на фотографиях или различные сложные применения вариационных квантовых схем, о которых пойдет речь далее в курсе.

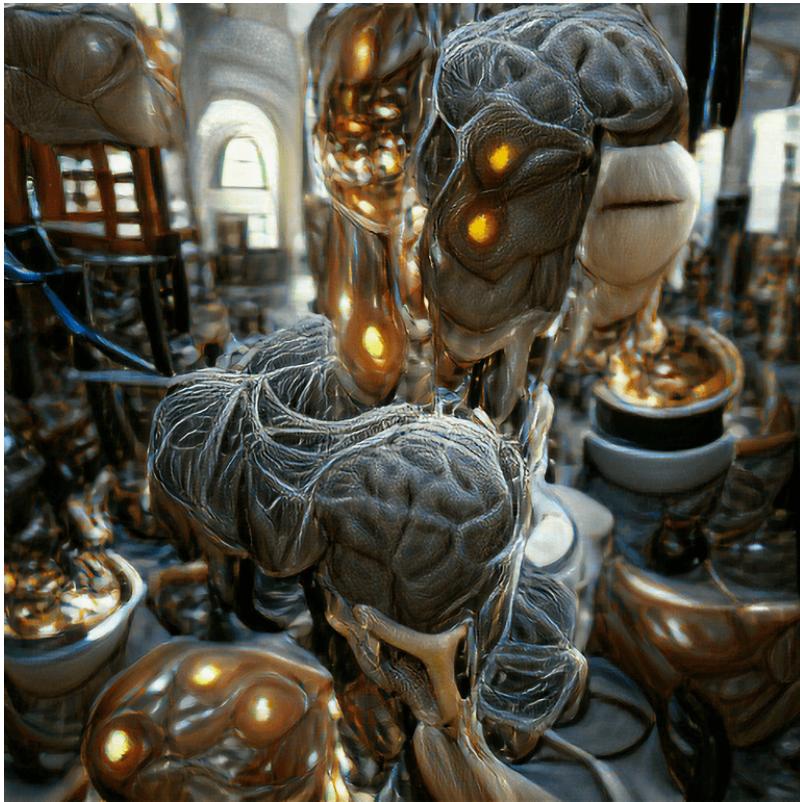


Fig. 42 Нейронные сети “глазами” других нейросетей – VQGAN и CLIP, сгенерировано в [Google Colab](#)

Машинное обучение как пазл

Введение

Мы продолжаем вводную часть нашего курса и переходим к машинному обучению. Если тема для вас хорошо знакома, можете пропустить этот блок, при желании можно пройти [этот тест](#), для определения достаточности уровня знаний (там 5 вопросов по Python, 5 – по машинному обучению и столько же – по линейной алгебре). Тест можно проигнорировать, если вы не знакомы с машинным обучением и для вас выглядят магией такие вещи как автоматическая детекция лиц на фото или определение тональности отзыва на товар.

Про машинное обучение, конечно, уже много всего написано, есть и немало неплохих курсов, сочетающих как теорию, так и практику. Но все же теория в этой области еще не догоняет практику, мы пока не понимаем, “почему оно работает”, а гарантии обобщающей способности алгоритмов (т.е. гарантии того, что модель машинного обучения будет работать на новых данных) в теории даются только для очень простых моделей. Таким образом, работа со сложными моделями остается своего рода искусством с примесью математики, инженерии и просто следования хорошим практикам, выработанным, как правило, в корпорациях или академическом сообществе.

В этой лекции мы примем сторону практики и расскажем про задачи машинного обучения как некоторый пазл (или лего, кому что ближе) – меняя разные кусочки, мы будем получать разные прикладные задачи/сценарии/модели применения машинного обучения. Для иллюстрации такое описание мы сопроводим 3-мя примерами:

- задача рекомендации контента и градиентный бустинг
- автоматическая оценка читаемости научной статьи и BERT
- детекция симптомов COVID-19 на рентгенограммах и YOLO

В этой лекции мы не опишем подробно, что это за модели машинного обучения (градиентный бустинг, BERT, YOLO), но зато покажем, что сценарии их применения в разных задачах (анализ табличных данных, текстов, изображений) похожи.

Note

Здесь мы почти не будем говорить о математике. Изложенный взгляд на машинное обучение как ремесло, вполне вероятно, вызовет критику со стороны специалистов в области статистики, эконометрики и теории машинного/статического обучения. Мы осознаем эти риски и тем не менее рассказываем о машинном обучении именно как о ремесле. Акцент в курсе делается на квантовые вычисления и квантовое машинное обучение, и в этой лекции мы опишем задачи "классического" машинного обучения на том уровне, чтобы просто было понятно, как это переносится на квантово-классические вариационные схемы и прочие алгоритмы, о которых пойдет речь далее в курсе. При этом строгость изложения материала тоже может немного пострадать.

Составляющие части задачи машинного обучения

Выделим следующие компоненты ("пазлы"), которые просматриваются во многих разных задачах машинного обучения:

- Целевой признак
- Модель
- Данные
- Функция потерь
- Решатель
- Схема валидации и метрика качества

По ходу изложения будем обсуждать упомянутые примеры задач машинного обучения.

Целевой признак

Есть задачи, в которых машинное обучение не нужно, а достаточно экспертных знаний. По закону Ома, известно что напряжение пропорционально силе тока и электрическому сопротивлению, и вряд ли захочется предсказывать напряжение в сети каким-то другим образом, кроме как применением закона Ома. То же самое можно сказать про многие другие физические явления.

Однако, для очень многих явлений вокруг нет хорошего теоретического объяснения или достаточных экспертных знаний. У нас нет "формулы", которая описала бы, как поставленный лайк к посту в соцсети, возврат кредита, клик по рекомендации товара или локализация заразы в конкретной части легких зависят от прочих факторов. В таких случаях мы можем приблизить такую неизвестную нам формулу с помощью машинного обучения.

В идеале с помощью машинного обучения мы хотели бы предсказывать какое-то событие, явление или процесс так, чтобы от этого была польза: прибыль компании/клиентов, если речь о бизнес-проекте, или новые знания, если это исследовательский проект. При этом напрямую это сделать вряд ли получится, и надо определить целевой признак, который, как мы считаем, будет связан с целевым событием/явлением. Звучит абстрактно, и дать строгое определение таких событий, явлений или процессов вряд ли получится. Поэтому сразу перейдем к примерам.

Note

Перед этим только небольшое замечание, что потребовав наличие целевого признака, мы ограничились рассмотрением задач обучения с учителем ([supervised learning](#)). Это все еще включает очень большой перечень типов задач машинного обучения, но не все.

Пример 1. Рекомендация новостного контента

Новостному порталу хочется понять, какой контент нравится пользователям и по каким ссылкам они будут кликать. Здесь *событием* будет то, что пользователю нравится рекомендуемый контент.



Fig. 43 Блок “Читайте также” на новостном портале

Понятно, что нет возможности установить строгую зависимость такого события от прочих факторов. Поэтому мы определяем целевой признак: факт клика пользователя по показанной рекомендованной ссылке. Мы верим, что клик по ссылке связан с *событием*: если пользователю нравится рекомендуемый контент, он/она перейдет по ссылке.

Пример 2. Автоматическая оценка читаемости научной статьи

Допустим, научному журналу хочется автоматически оценивать читаемость текста, чтобы знать, какие статьи можно сразу подавать на рецензию, а какие лучше предварительно направить в сервис proofreading, где статья будет вычитана и поправлена носителем языка.

В идеале мы бы хотели предсказывать, “хорошо” ли написана статья или “плохо”. Но это очень сложно определить формально, и потому есть много метрик читаемости текста, таких как [Automated readability index](#) или [Flesch reading ease](#), которые являются эвристиками и “приближают” то что мы имеем в виду под “хорошо” или “плохо” написанным текстом. Кстати, на момент написания это лекции на платформе Kaggle проходит [соревнование](#) по этой теме.

Небольшое лирическое отступление: во многом опыт специалиста по машинному обучению оказывается в способности понять, когда это машинное обучение **не** требуется. Описанную задачу можно решить и без всякого машинного обучения. Можно замерить 5-10 метрик читаемости текста, разметить 100-200 статей вручную (желательно, чтоб это делали эксперты уровня редактора журнала, а еще и лучше бы иметь по 3 оценки на статью) и заключить, хорошо ли метрики читаемости коррелируют с оценками экспертов. Другой вариант – попытаться малой ценой использовать готовые решения, например, Grammarly.

Но если этого окажется недостаточно, придется подумать. Пока остановимся тут и еще раз подчеркнем, что определить читаемость текста напрямую – невозможно, и мы это заменим на другой Целевой признак, например, на агрегированную метрику читаемости текста.

Пример 3. Детекция симптомов COVID-19 на рентгенограммах

Последние пару лет мы видели бурное развитие методов глубокого обучения в приложениях к анализу медицинских данных, а в особенности это стало актуальным в симптомах COVID-эпохи.

Допустим, стоит задача определения аномалий на рентгенограммах грудной клетки. В идеале мы хотели мы сразу по таким изображениям обнаруживать симптомы симптомов COVID-19 у пациента. Но заголовок этого примера выдает желаемое за действительное, и, конечно, сразу по снимкам диагностировать не получится.



Fig. 44 Пример данных [соревнования](#) по определению аномалий на рентгенограммах грудной клетки.

[Источник](#)

Поэтому в такой задаче надо аккуратно определить *Целевой признак*. В данном случае их несколько. Согласно [описанию источника данных](#) соревнования, комитет врачей-радиологов проанализировал 6334 рентгенограмм и пометил их метками: 'Negative for Pneumonia' (нет пневмонии), 'Typical Appearance' (нормально), 'Indeterminate Appearance' (неразборчиво) и 'Atypical Appearance' (ненормально). Надо четко понимать, что возможности обученной модели будут ограничены имеющейся разметкой и поэтому заголовок "детекция симптомов COVID-19" слегка "желтоват", в реальности модель детекции сможет выделять участок изображения (bounding box) и помечать это вектором из 4-х значений, соответствующих описанным целевым признаком в обучающей выборке.

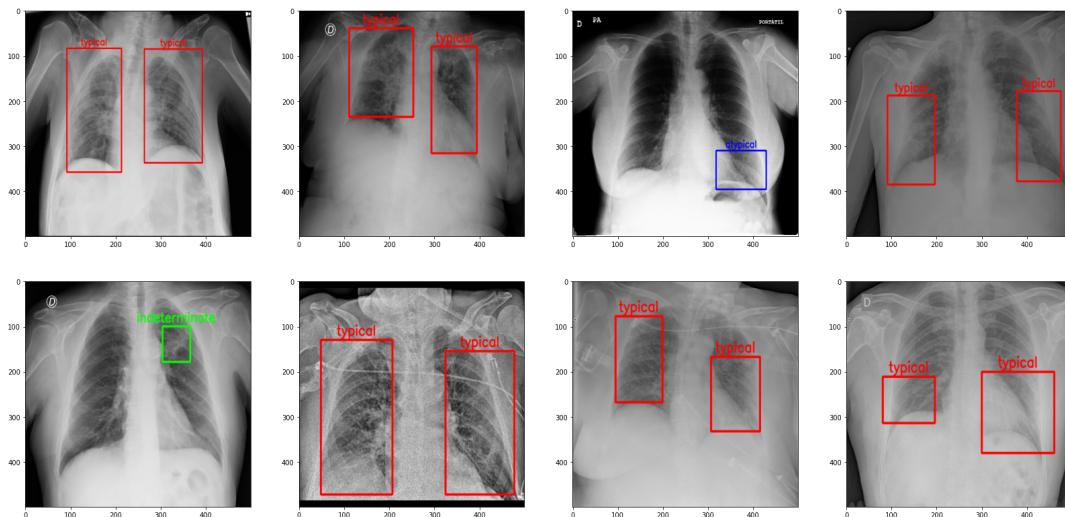


Fig. 45 Прогноз модели детекции, обученной на данных [соревнования](#) по определению аномалий на рентгенограммах грудной клетки. [Источник](#)

Данные

Определение *Целевого признака* неразрывно связано с имеющимися данными. Нет смысла определять целевой признак, который мы не можем измерить или по которому мы не можем собрать данные. Например, если новостной портал не логирует клики пользователей, нет смысла задавать вопрос о том, нравятся ли пользователям показанные рекомендации – для начала надо настроить сервисы для хранения кликов. Другой пример: вряд ли стоит пытаться предсказывать движения денежных активов в микросекундном диапазоне, если нет дорогостоящей инфраструктуры для сбора и обработки таких данных.

Но *Данные* – это, конечно, не только *Целевой признак*. Но и просто *признаки*, также в эконометрике называемые предикторами или независимыми переменными. С помощью *признаков* как раз и получится предсказать *Целевой признак*, а хорошо или плохо – об этом чуть далее.

Продумывание, сбор, и обработка *признаков* – один из наиболее творческих аспектов работы специалистов по машинному обучению. Эта работа может включать в себя общение с представителями предметной области или бизнеса, клиентами, чтение научных статей, применение и более технических приемов, таких как порождение

признаков из других признаков (feature engineering) или использование моделей для получения признаков (например, признаки изображения, полученные предобученной искусственной нейросетью).

Пример 1. Рекомендация новостного контента

В этой задаче, как правило, хорошо работают данные о поведении пользователей, попросту – “клики”. Обычно алгоритмы рекомендации хорошо работают при обучении на больших объемах данных, и поэтому большая часть проекта посвящена настройка хранилища данных и обработке потоковых событий (показ рекомендации, клик по рекомендации).

Если есть возможность собрать какие-то признаки пользователей (пол, возраст, указанные интересы) или рекомендуемого контента (темы новостей, представление текста новости “эмбеддингами”), эту информацию можно добавить в модель.

При достаточно хорошей подготовке данных задачу можно решить и без машинного обучения. Довольно сильным прототипом (baseline-решением), опять же, при достаточных объемах данных, может быть просто сортировка контента по Click-through-Rate (CTR). Для каждой новости мы просто собираем статистику, сколько раз она была показана пользователю в качестве рекомендации и сколько раз по ней в итоге кликнули. Отношение кликов к показам и есть Click-through-Rate. Есть, конечно, детали – что делать с “холодными” новостями, без накопленной статистики для расчета CTR, что делать с clickbait-новостями, как фильтровать то, что нельзя показывать в качестве рекомендации. Но после решения этих проблем может оказаться, что просто сортировка по CTR – уже неплохое решение.

Пример 2. Автоматическая оценка читаемости научной статьи

Допустим, научному журналу удалось договориться с сервисом proofreading и получить данные о том, какие статьи хорошо написаны и не требуют множества правок, а какие пришлось переписывать почти что с нуля. Эту информацию можно пытаться использовать для обучения модели, которая для заданного куска текста будет предсказывать, как сильно его нужно поправить.

Тут мы будем иметь дело с задачей из области NLP (Natural Language Processing) – на входе будет просто текст (полный текст статьи или разбитый на параграфы) и по сути мы можем не придумывать признаки вручную, а модель сама их извлечет, то есть выучит представление текста. Целевым признаком при этом будет, например, какое-либо расстояние (скажем, Левенштейна) между оригинальным текстом и поправленным редактором. Таким образом, это будет задачей регрессии, в которой для текста предсказывается, как сильно его следует изменить.

Опять стоит отметить, что в такой задаче скорее всего нужно много обучающих данных, чтобы описанный алгоритм хорошо заработал.

И еще стоит отметить, что задачи NLP зачастую пересекаются с лингвистикой, и это как раз такой пример. Вполне вероятно, что задачу можно решить без тоже машинного обучения и по-другому – на основе правил-эвристик, разработанных в сотрудничестве с лингвистами. Эдакая версия Grammarly для работы с научными текстами.

Но дальше в примере будем считать, что лингвистов в команде нет, простые метрики читаемости текста, описанные выше, работают плохо, и мы решаем задачу регрессии, то есть используем машинное обучение.

Пример 3. Детекция симптомов COVID-19 на рентгенограммах

В этой задаче ключевые данные для обучения модели – это собственно изображение, рентгенограмма области грудной клетки и разметка, которая состоит из координат интересующей области изображения и типа области, в данном случае это одна из 4-х меток: ‘Negative for Pneumonia’ (нет пневмонии), ‘Typical Appearance’ (нормально), ‘Indeterminate Appearance’ (неразборчиво) и ‘Atypical Appearance’ (ненормально).

Конечно, у снимков есть разные метаданные, да и форматы медицинских данных обычно специфические, но нам для примера подойдет такое упрощение.

Модель

С моделированием знаком любой исследователь. Чтобы рассчитать минимальную толщину стекла вагона метро для защиты от птиц во время движения по открытым участкам, достаточно представить птицу цилиндром той же массы, и для данной задачи цилиндр будет подходящей моделью птицы.

Подобным же образом, в задачах машинного обучения с учителем Модель приближает Целевой признак и делает это с помощью *Данных и Параметров*. (Параметры – это неотъемлемая часть модели, и поэтому мы их не выносим как отдельный компонент задач машинного обучения).

Надо понимать, что предлагая модель, мы совершаём уже второе упрощение. Сначала, как мы говорили, Целевой признак заменяет нам то, что мы реально хотим знать. А теперь, к тому же, мы заменяем целевой признак на его прогноз с помощью модели.

Пример 1. Рекомендация новостного контента

В задачах рекомендации есть классический алгоритм ALS (Alternative Least Squares), но можно задачу решить и как задачу ранжирования. Это может быть предпочтительно, поскольку можно использовать боевую лошадку машинного обучения на табличных данных – градиентный бустинг (доступно, на русском про эту модель написано [тут на Хабре](#) в рамках курса [mlcourse.ai](#)). Бустинг подходит для задач классификации, регрессии и ранжирования, и его можно использовать также и в описанной задаче.

Также, если бустинг уже используется в компании в других задачах, скорее всего получится безболезненно переиспользовать опыт поддержки модели и соответствующей инфраструктуры в "продакшене" вместо того, чтобы отдельно все это разрабатывать для ALS.

Пример 2. Автоматическая оценка читаемости научной статьи

Описанную задачу, опять же с оговорками про возможность альтернативного подхода без всякого машинного обучения, скорее всего хочется решать с помощью языковых моделей, основанных на трансформерах. В частности, в этой задаче имеет смысл использовать [SciBERT](#), предобученный как раз на научных текстах. Это модель типа BERT (Bidirectional Encoder Representations from Transformers, [arXiv](#)), которая основана на архитектуре трансформеров, ставшей настоящей революцией в NLP (оригинальная статья – [Attention is all you need](#), NIPS 2017). BERT используется для представления текста на основе трансформеров и дообучения параметров под многие стандартные задачи NLP – классификация текстов, пар текстов, вопросно-ответные системы, распознавание именованных сущностей и т.д. Подробнее про BERT можно почитать в [постах Jay Alammar](#) (англ.) или в их переводах на русский: ["Transformer в картинках"](#) и ["Ваш первый BERT: иллюстрированное руководство"](#).

Пример 3. Детекция симптомов COVID-19 на рентгенограммах

Подходов к детекции объектов на изображениях немало, но по соотношению скорости и качества работы особенно хорошо себя зарекомендовала модель [YOLOv5](#). Про принцип работы этой модели можно узнать из лекции ["Detection and segmentation"](#) курса cs231n, также на Хабре можно найти [статью](#) про YOLOv4.



Fig. 46 Пример обнаружения объектов на изображении. [Источник](#)

Функция потерь

Выбор функции потерь (loss function) зависит от конкретной задачи, и это вопрос, изучаемый в курсах машинного обучения. Функция потерь определена для объектов обучающей выборки и по сути говорит, насколько прогноз хорошо соответствует значению целевого признака.

Тут тонкий момент: примерно для того же нужны метрики качества, о которых речь пойдет ниже. Но функция потерь на практике чаще всего используется именно для того, чтобы задать цель обучения модели (для чего именно ей менять свои параметры) и также оценить, насколько хорошо модель обучилась, попросту, насколько хорошо она “сочлаась”.

В отличие от метрик качества, функции потерь вполне могут быть плохо интерпретируемыми, например как логистическая функция потерь (logloss, на русском про нее можно прочитать в [посте](#) А.Г. Дьяконова), и на практике при разработке модели Data Scientist посмотрит на значение функции потерь всего несколько раз:

- при отладке модели стоит проверить, может ли она “переобучиться под мини-батч”, то есть может ли она при обучении всего с парой десятком примеров добиться почти нулевого значения функции потерь. Это важно, чтобы понять, нет ли где-то ошибки в коде описания модели и хватает ли модели сложности (capacity), чтобы подстроиться под данные
- чтобы избежать переобучения, стоит проверять (вручную или автоматически) значение функции потерь на отложенной выборке
- еще значения функции потерь можно сравнивать для разных версий модели, чтобы понять, какая из них лучше обучилась

Заметим, что дизайн функции потерь под задачу, как и придумывание признаков – порой интересный творческий процесс, а итоговая функция потерь, используемая для обучения модели может быть сложной, состоящей из нескольких более простых функций потерь.

Для примера, в задаче переноса стиля (style transfer), в классическом варианте задаются два изображения – “контентное” и “стилевое” – и генерируется третье изображение, которое похоже в целом на “контентное” изображение, но по стилю – на “стилевое”. При этом функция потерь складывается из двух других:

- одна – content loss – передает, насколько отличаются карты признаков (feature maps) генерируемого и “контентного” изображений
- вторая – style loss – соответственно передает, насколько похожи стили генерируемого и “стилевого” изображений. Делается это хитро, и за деталями лучше обратиться, например, к [лекции “Visualizing and Understanding”](#) стэнфордского курса cs231n.



Fig. 47 Пример решения задачи Neural Style Transfer из [задания 3](#) стэнфордского курса cs231n.

Функция потерь может включать и много составляющих, больше двух, если мы хотим чтобы модель выучила разные аспекты задачи. К примеру, в [этой статье на Хабре](#) (уровень - продвинутый NLP) Давид Дэлл описывает дистилляцию нескольких больших NLP моделей для получения маленькой версии русскоязычной модели BERT. Маленький BERT по сути “учится” у больших моделей RuBERT, LaBSE, USE и T5, а описание того, что маленькая модель должна уметь (предсказывать замаскированные токены по контексту, строить представления токенов подобно тому, как это делают большие модели, предсказывать правильный порядок токенов в предложении) – это и есть составление сложной функции потерь.

Пример 1. Рекомендация новостного контента

Задача свелась к задаче ранжирования на табличных данных, и тут можно использовать функцию потерь, которую можно оптимизировать с помощью градиентного бустинга (т.е. дифференцируемую, это важно), для задачи ранжирования. Например, [LambdaMART](#).

Пример 2. Автоматическая оценка читаемости научной статьи

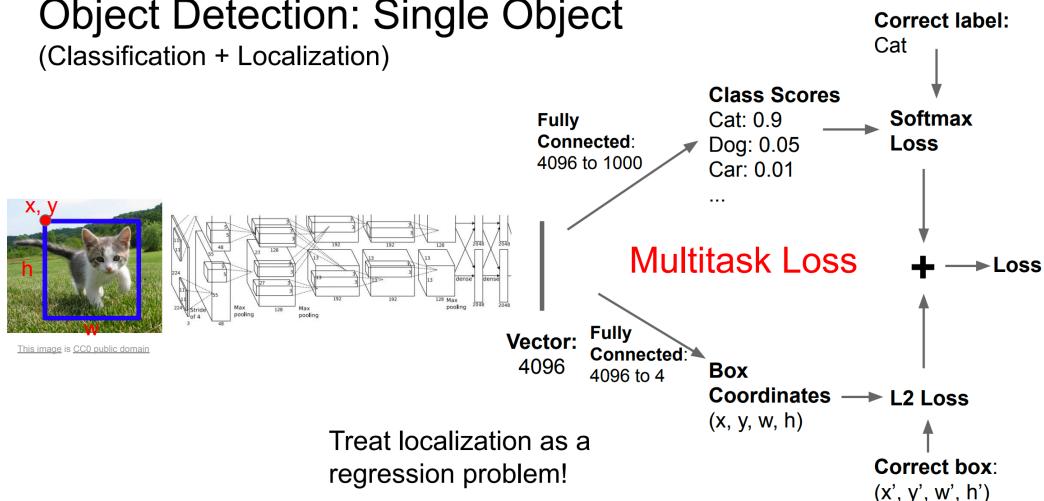
Тут задача свелась к задаче регрессии с текстовыми входными данными, и можно использовать простую функцию потерь – среднеквадратичную ошибку (Mean Squared Error). Подробнее о функциях потерь в задаче регрессии можно прочитать в [посте](#) А.Г. Дьяконова.

Пример 3. Детекция симптомов COVID-19 на рентгенограммах

В задаче детекции объектов на изображениях, как правило, для каждого объекта модель выдает 4 числа – координаты окна (bounding box) и его длину и ширину, а также вектор с числами – оценки принадлежности данного региона к каждому из классов. Поэтому функция потерь, как правило, складывается из двух других:

- Для сравнения предсказанного региона с реальным (согласно разметке в обучающей выборке) используется среднеквадратичная ошибка (Mean Squared Error).
 - Для сравнения вектора оценок принадлежности региона к каждому из классов с реальным (согласно разметке в обучающей выборке) используется стандартная для задачи классификации функция потерь – логистическая, (также известная под термином “кросс-энтропия”), которую мы упоминали выше.

Object Detection: Single Object (Classification + Localization)



Fei-Fei Li, Ranjay Krishna, Danfei Xu

Lecture 12 - 41

May 19, 2020

Fig. 48 Слайд стэнфордского курса cs231n про функцию потерь в задаче детекции объектов.

Источник

Решатель

Когда мы определились с целевым признаком, обучающими данными, моделью и функцией потерь, мы наконец можем собственно обучать модель – то есть менять ее параметры так, чтобы все лучше и лучше предсказывать целевой признак, что меряется с помощью функции потерь.

Наиболее часто используемых вариантов два:

- явное решение (closed-form solution), в котором просто применится формула для получения оптимальных параметров модели. Тут Решателем будет фреймворк, вычисляющий эту явно выписанную формулу, т.е. скорее всего производящий матричные умножения – например, NumPy.
 - оптимизация параметров модели градиентными методами оптимизации. Тут Решателем будет алгоритм оптимизации и его конкретная реализация, например, в PyTorch

Самый известный пример первого варианта – это явное решение задачи наименьших квадратов. Существует прямая конкретная формула, включающая в себя перемножение матриц и векторов и взятие обратной матрицы, и дающая оптимальный (по минимизации среднеквадратичной ошибки) вектор весов линейной модели при наличии, конечно, обучающей выборки.

Получилось, что эта лекция вводная, без математических выкладок, поэтому не будем здесь приводить формулы без контекста, детали можно найти в любом классическом учебнике по статистике или машинному обучению или вкратце [в этой статье](#) на Хабре в разделе “Линейная регрессия”.

Во всех трех примерах задач, которые мы попутно рассматриваем и которые сводятся к обучению градиентного бустинга, SciBERT и YOLO соответственно, нет явного решения, которое простой формулой выдало бы оптимальные параметры модели, поэтому параметры подбираются с помощью методов оптимизации. В примере с градиентным бустингом “под капотом” – алгоритм обучения деревьев решений, с ним можно познакомиться в [этой статье](#). А в двух других примерах обучаются искусственные нейронные сети, и сегодня это делается как правило с помощью оптимизаторов, реализованных во фреймворках (PyTorch/TensorFlow/etc), причем аппаратная реализация, в которой работает *Решатель*, – это скорее видеокарта или ТПУ.

Ко второму из рассмотренных вариантов Решателя можно также отнести и автоматическое дифференцирование. Чтобы оптимизировать параметры модели градиентными методами, нужно знать собственно градиенты функции потерь – вектора частных производных функции потерь по параметрам. Попросту говоря, это ответ на вопрос, какие параметры модели стоит изменять и как это скажется на функции потерь, которую мы хотим минимизировать.

Иногда такие градиенты функции потерь можно найти аналитически, например, в случае линейной или логистической регрессии. Но как правило, в случае более сложных моделей с большим числом параметров аналитическое выражение для градиента функции потерь выглядит слишком громоздко, лучше вычислять его численно. Это задача методов автоматического дифференцирования. В частности, возможно, величайший алгоритм машинного обучения всех времен – алгоритм обратного распространения ошибки (backpropagation) – не что иное как численный метод нахождения производных функции потерь по параметрам модели.

Схема валидации и метрика качества

Наконец, когда мы определились с целевым признаком, обучающими данными, моделью, функцией потерь, а также научились подбирать параметры модели так, чтобы функция потерь уменьшалась, остался последний шаг – решить, как мы действительно поймем, что решаем задачу хорошо.

Выбор метрики качества напрямую связан с тем, чего мы хотим от модели машинного обучения в более широком контексте. Например, в бизнес-процессе могут быть ключевые показатели, которые мы *опосредованно* можем улучшать с помощью машинного обучения. Ключевыми показателями могут быть такие вещи как дневная аудитория приложения, Life-Time Value, показатели, связанные с удержанием (retention) клиентов/сотрудников, удовлетворенность клиентов и т.д. Многие из этих показателей нельзя замерять напрямую и оптимизировать, и тогда искусство Data Scientist-а заключается в том, чтобы выбрать простую метрику качества, которая бы задавалась понятной формулой (например, доля верных ответов или полнота) и при этом неплохо “коррелировала” с тем показателем бизнес-процесса, который хочется оптимизировать и таким образом приносить деньги компании или уменьшать операционные расходы.

Определение схемы валидации нужно, чтобы ответить на следующие вопросы:

- как понять, что модель сработает неплохо на новых, ранее не виденных данных
- как понять, что мы улучшили решение, поменяв модель, ее гиперпараметры или добавив новые признаки
- как понять, что одна модель лучше другой модели, один набор признаков лучше другого при фиксированной модели и т.д.

Во многом это связано с переобучением. Недостаточно просто замерить метрику качества на обучающей выборке. Надо хотя бы разбить выборку на 2 части: на одной обучать модель, на второй – проверить метрику качества. И чаще всего при больших объемах данных и больших моделях ровно так и делают. Но в мире “малых данных” и легковесных моделей более предпочтительна кросс-валидация. В этой схеме выборка делится на несколько частей, а модель обучается столько же раз. При этом каждая из подвыборок один раз является тестовой частью, на которой измеряется качество прогнозов, а все остальные разы она участвует в обучении модели. Таким образом, кросс-валидация дает более надежную оценку того, как модель сработает на новых данных, в сравнении с простым разбиением обучающей выборки на две части.

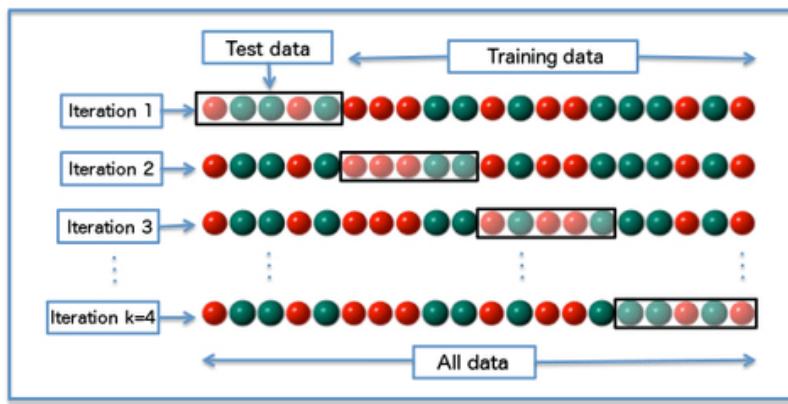


Fig. 49 Схема K-fold кросс-валидации.

Вы вряд ли ошибетесь, если на практике будете применять 5-кратную стратифицированную (такую что распределение целевого признака примерно одинаковое в каждой подвыборке) кросс-валидацию, но надо понимать, что кросс-валидация – не панацея. И схема валидации, которую мы тут описываем, – это больше, чем

просто кросс-валидация.

Часто выделяют “оффлайн” и “онлайн” режимы валидации. И метрики, полученные на кросс-валидации относятся к первому режиму, “оффлайн”. То есть мы один раз собрали обучающие данные, они больше не меняются, и вот с таким статическим срезом данных мы работаем, создаем признаки, обучаем модели, измеряем значения метрик на кросс-валидации. Но это не отвечает на вопрос, что получится, когда мы “выкатим модель в прод”, какие значения метрик ожидать на новых данных. И тут, как правило, устраиваются A/B-тесты, которые позволяют понять, а действительно ли мы видим эффект от модели, лучше ли вообще с моделью, чем без нее, а если лучше с моделью, то какую модель выбрать. A/B-тестирование – это очень обширная тема, выходящая за рамки данной статьи, и развивать ее тут мы не будем. Только отметим, что определение схемы валидации – зачастую нетривиальный процесс и в зависимости от проекта может быть методологически сложным, в том числе и приводить к ошибкам.

Также заметим, что AutoML, о котором столь многие мечтают, решает некоторые вопросы, но далеко не все. В частности, алгоритмы AutoML не подберут вам правильную схему валидации. Они работают с уже имеющейся схемой валидации, и если валидация неверна, приводят к ошибкам, то тут и AutoML не поможет. Так что Data-Scientist-ов AutoML пока не вытеснит.

Пример 1. Рекомендация новостного контента

В этой задаче для выбора модели и ее гиперпараметров можно использовать обычную кросс-валидацию, но вот чтобы убедиться, что “в бою” модель также работает, лучше настроить A/B-тест (а точнее, для задач ранжирования намного лучше использовать [интерливинг](#)). Пришло время сказать, что этот пример взят из практики автора этой лекции, Юрия Кашницкого, и в [выступлении на DataFest 2018](#) описывались сложности валидации модели в задаче рекомендации новостей. Вывод такой, что в задачах рекомендации (да и во многих других) надо устроить онлайн-проверку модели (A/B-тест, интерливинг) помимо оффлайн-проверки (кросс-валидации), только таким образом можно удостовериться, что модель действительно полезна, например, что использовать модель – лучше, чем просто показывать самый популярный контент.

Пример 2. Автоматическая оценка читаемости научной статьи

Здесь тоже может использоваться обычная кросс-валидация, хотя на практике из-за объемов данных и размера модели кросс-валидацию проводить будет дорогостоящим и придется удовлетвориться разбиением обучающей выборки на две части и проверкой модели на отложенной части.

Есть, конечно, детали. BERT не очень хорошо работает с длинными текстами, так что скорее всего мы разобьем полный текст статьи на параграфы и будем их подавать в модель по очереди. В таком случае лучше проводить [GroupKFold кросс-валидацию](#) так, чтобы на каждом этапе кросс-валидации в обучающей и проверочной выборке были параграфы из разных статей. Мы вряд ли хотим обучаться на одной половине статьи и проверять модель на второй ее половине – так бы мы получили слишком оптимистичную оценку качества модели.

Но даже при оговоренных тонкостях кросс-валидации самая большая сложность данной задачи – убедиться, что метрика качества соответствует тому, что мы реально хотим получить в задаче. Это связано со сложностью определения “хорошо” и “плохо” написанного научного текста, о которой мы говорили выше. Поэтому в данной задаче, как и во многих других практических задачах машинного обучения, не обойтись без проверки результатов модели вручную. Такая проверка модели будет делаться уже после кросс-валидации, и в этом смысле она похожа на онлайн-оценку модели.

Пример 3. Детекция симптомов COVID-19 на рентгенограммах

Здесь схема проверки модели очень похожа на предыдущую. Сначала кросс-валидация или проверка модели на отложенной части, а потом – проверка предсказаний модели экспертами. Конечно, никто не будет по одной только кросс-валидации или результатам участников в соревновании Kaggle заключать, что модель прекрасно работает и ее можно нести врачам. В данном случае надо проверить модель на данных, приближенных к тем, которые будут использоваться врачами на практике. Чтобы не было [таких историй](#), как у Google Health, когда Deep Learning модель достигала 90% верных ответов при определении диабетической ретинопатии по фото зрачка, но при обучении на качественных снимках высокого разрешения. А при работе с менее качественными снимками система просто слишком часто отказывалась выносить вердикт из-за того, что была недостаточно уверена в прогнозе.

Заключение

В этой лекции мы описали, из чего складывается постановка задачи машинного обучения и рассмотрели, как общие компоненты проглядываются в разных по своей природе задачах. При этом мы поговорили о моделях-рабочих лошадках в трех разных областях: градиентном бустинге для табличных данных, BERT для текстов и YOLO для детекции изображений.

Немного пожертвовав, возможно, строгостью определения таких понятий как целевой признак или решатель, мы, надеюсь, описали все "пазлы" достаточно абстрактно, чтобы сложилось общее представление о том, как машинное обучение применяется в разных задачах, а также какие подводные камни стоит ожидать при боевом применении машинного обучения. Надеемся, это позволит лучше осознать взаимосвязь разных компонентов в квантово-классических схемах обучения, о которых речь пойдет далее в курсе.

О блоке “Квантовые вычисления”

Этот блок включает в себя:

- общий рассказ о том, что такое квантовый бит;
- введение в основные квантовые гейты.

Продвинутые темы блока дополнительно рассказывают:

- о квантовой физике, на которой базируется концепция кубита;
- о смешанных состояниях, операторе плотности и энтропии фон Неймана.

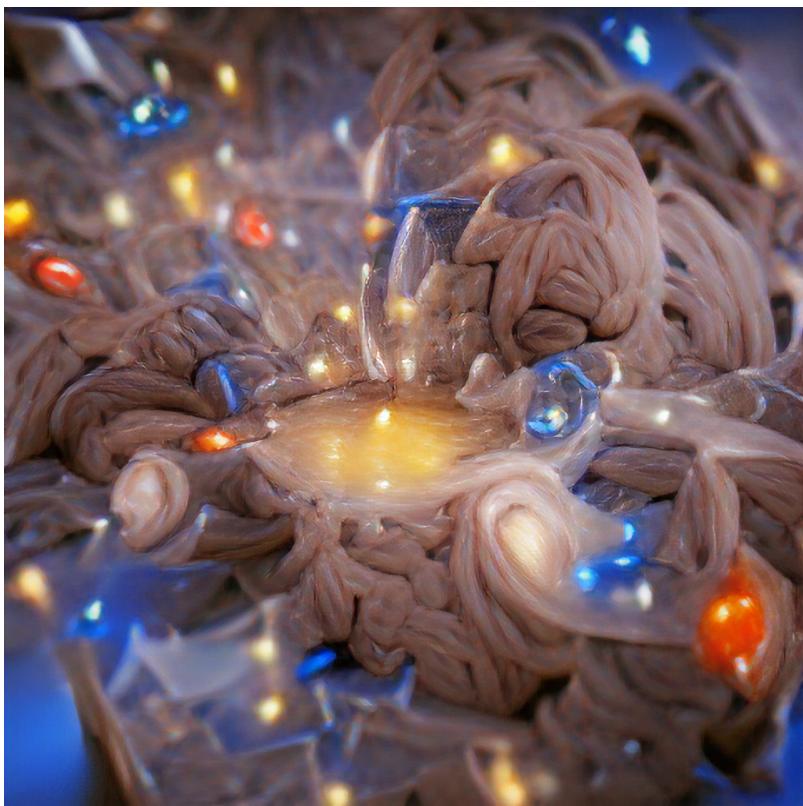


Fig. 50 Квантовый разум "глазами" VQGAN и CLIP, сгенерировано в [Google Colab](#)

Квантовый бит

Описание лекции

Эта лекция расскажет:

- что такое кубит;
- в чем разница между значением и состоянием;
- что такое сфера Блоха;
- какие можно делать операции над кубитами;

- что такое измерение.

Введение

Это первая лекция основного блока нашего курса. Прежде чем мы начнем детально разбирать понятие кубита, давайте взглянем на общий пайплайн квантовых схем.

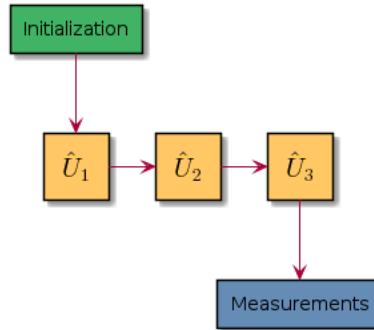


Fig. 51 Схема любого квантового алгоритма

Любая квантовая схема включает в себя:

- кубиты, инициализируемые в начальное состояние, обычно $\lvert \text{ket}\{0\} \rangle$;
- унитарные и обратимые операции над кубитами;
- измерение кубитов.

Эта лекция посвящена разбору операций для одного кубита. Начнем с понятия кубита и его отличий от бита классических компьютеров.

Что такое кубит

Классический компьютер оперирует двоичными числами – нулем и единицей. Минимальный объем информации для классического компьютера называется бит. Квантовый компьютер оперирует квантовыми битами или кубитами, которые тоже имеют два возможных значения – 0 и 1. Так в чем же разница? В чем особенности квантовых компьютеров, которые дают им преимущества над классическими компьютерами?

Разница в том, что для квантомеханических систем (и кубитов в частности) их состояния и значения – это не одно и то же.

Состояние vs значение

Состояние классического бита

Обычно мы не отличаем состояние классического бита от его значения и считаем, что если бит имеет значение **1**, то и состояние его описывается числом **1**.

Кот Шредингера

Давайте вспомним мысленный эксперимента Шредингера. Кот, который одновременно и жив, и мертв. Понятно, что значение кота точно одно: он либо жив, либо мертв. Но *состояние* его более сложное. Он находится в *суперпозиции* состояний “жив” и “мертв” одновременно.



Fig. 52 Кот Шредингера

Состояние кубита

Состояние кубита, если можно так сказать, аналогично состоянию кота Шредингера. Оно отличается от значения кубита и описывается вектором из двух комплексных чисел. Мы будем обозначать состояния (или вектора) символом $\langle \Psi |$ (кет – вектор-столбец) – это широко принятая в квантовой механике и квантовых вычислениях нотация Дирака:

$\begin{aligned} \begin{array}{l} \text{ket} \end{array} \Psi = \begin{array}{l} \text{bmatrix} \end{array} c_0 \\ c_1 \end{array} \end{aligned}$

Note

Может возникнуть вопрос, а почему комплексные числа? Короткий ответ на этот вопрос дать сложно. Если в двух словах, то использование комплексных чисел связано с удобством представления матричных групп, используемых в квантовой механике.

Все еще звучит сложно? Тогда нужно вспомнить, что изначально квантовая механика возникла в том числе из-за того, что физики экспериментально обнаружили у фундаментальных частиц свойство **корпускулярно-волнового дуализма**. Иными словами, электроны,光子ы и другие частицы проявляли как типичные свойства волнового движения (например, интерференцию и дифракцию), и свойства частиц – например, всегда есть минимальная порция (**квант!**) света или электрического поля. Кстати, часто вместо вектора состояния используется понятие **волновой функции**, которая описывает плотность вероятности обнаружить частицу в той или иной точке пространства (обычного или специального). Ко времени создания квантовой механики для описания волнового движения ученые уже привыкли использовать комплексные числа, которые позволяют упростить описание многих эффектов за счет разделения амплитуды и фазы процесса. Такое удобство справедливо и для многих задач квантовой физики.

Для более детального ответа авторы курса рекомендуют читать книги по истории квантовой физики (и по самой квантовой физике).

Значение чисел $|c_0\rangle$ и $|c_1\rangle$ мы обсудим чуть позже, а пока запишем наш кубит $\langle \Psi |$ в коде Python. Для начала $c_0 = c_1 = \frac{1}{\sqrt{2}}$.

```
import numpy as np
qubit = np.array([1 / np.sqrt(2) + 0j, 1 / np.sqrt(2) + 0j]).reshape((2, 1))
```

Здесь мы создаем именно вектор-столбец размерности (2×1) .

```
print(qubit.shape)
```

```
(2, 1)
```

Связь состояния и значения кубита

Разберем подробнее вектор $\langle \Psi |$ и значение цифр $|c_0, c_1\rangle$. Посмотрим на состояния кубита, значение которого мы знаем точно. То есть “посмотрим на кота Шредингера”, но который точно жив или точно мертв.

Базисные состояния

Посмотрим, как выглядят состояния кубитов с точно определенными значениями:

```
\begin{split} \ket{0} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{quad} \ket{1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{split}
```

Что мы можем сказать об этих состояниях? Как минимум следующее:

- они ортогональны ($(\langle \ket{0} | \perp \rangle \langle \ket{1} |)$);
- они имеют единичную норму;
- они образуют базис.

Что это значит для нас? А то, что любое состояние $|\Psi\rangle$ можно записать как линейную комбинацию векторов $|\ket{0}\rangle$ и $|\ket{1}\rangle$, причем коэффициентами в этой комбинации будут как раз наши (c_0, c_1) :

```
basis_0 = np.array([1 + 0j, 0 + 0j]).reshape((2, 1))
basis_1 = np.array([0 + 0j, 1 + 0j]).reshape((2, 1))

c0 = c1 = 1 / np.sqrt(2)

print(np.allclose(qubit, c0 * basis_0 + c1 * basis_1))
```

True

Амплитуды вероятностей

Квантовая механика устроена таким интересным образом, что если мы будем измерять **значение** кубита, то вероятность каждого из вариантов будет пропорциональна соответствующему коэффициенту в разложении **состояния**. Но так как амплитуды – это в общем случае комплексные числа, а вероятности должны быть строго действительные, нужно домножить амплитуды на комплексно сопряженные значения. В случае наших значений $(c_0 = c_1 = \frac{1}{\sqrt{2}})$ получаем:

```
p0 = np.conj(c0) * c0
p1 = np.conj(c1) * c1

print(np.allclose(p0, p1))
print(np.allclose(p0 + p1, 1.0))
```

True
True

Видим еще одну важную вещь: сумма вероятностей всех состояний должна быть равна 100%. Это сразу приводит нас к тому, что состояния – это не любые комплексные вектора, а комплексные вектора с единичной нормой:

```
print(np.allclose(np.conj(qubit).T @ qubit, 1.0))
```

True

Мы будем очень часто пользоваться транспонированием и взятием комплексно сопряженного от векторов. В квантовой механике это имеет специальное обозначение $|\bra{\Psi}| = |\Psi^\dagger|^*$ ($\bra{\Psi}$ – бра – вектор-строка). Тогда наше правило нормировки из **NumPy** кода может быть записано в нотации Дирака так:
 $|\braket{\Psi | \Psi}| = 1$

Сфера Блоха

Описанный выше базис $|\ket{0}, |\ket{1}\rangle$ не является единственным возможным. Вектора $|\ket{0}, |\ket{1}\rangle$ – это лишь самый часто применимый базис, который называют $|\mathbf{Z}\rangle$ базисом. Но есть и другие варианты.

Возможные базисы

Z-базис

Уже описанные нами $|\ket{0}\rangle$ и $|\ket{1}\rangle$.

X-базис

Базисные состояния $\lvert \text{ket}{+} \rangle = \frac{\lvert \text{ket}{0} \rangle + \lvert \text{ket}{1} \rangle}{\sqrt{2}}$ и $\lvert \text{ket}{-} \rangle = \frac{\lvert \text{ket}{0} \rangle - \lvert \text{ket}{1} \rangle}{\sqrt{2}}$:

```
plus = (basis_0 + basis_1) / np.sqrt(2)
minus = (basis_0 - basis_1) / np.sqrt(2)
```

Y-базис

Базисные состояния $\lvert \text{ket}{R} \rangle = \frac{\lvert \text{ket}{0} \rangle + i\lvert \text{ket}{1} \rangle}{\sqrt{2}}$ и $\lvert \text{ket}{L} \rangle = \frac{\lvert \text{ket}{0} \rangle - i\lvert \text{ket}{1} \rangle}{\sqrt{2}}$:

```
R = (basis_0 + 1j * basis_1) / np.sqrt(2)
L = (basis_0 - 1j * basis_1) / np.sqrt(2)
```

Легко убедиться, что все вектора каждого из этих базисов ортогональны:

```
print(np.allclose(np.conj(basis_0).T @ basis_1, 0))
print(np.allclose(np.conj(plus).T @ minus, 0))
print(np.allclose(np.conj(R).T @ L, 0))
```

```
True
True
True
```

Заметьте, что в наших векторных пространствах скалярное произведение – это (в случае действительных векторов) $\langle \text{vec}{a} | \text{vec}{b} \rangle = \left| \begin{array}{c} \text{vec}{a} \\ \text{vec}{b} \end{array} \right|$ ([бра-кет](#)). Именно поэтому нужно делать транспонирование и комплексное сопряжение первого вектора в паре.

Сфера Блоха

Обозначения $\lvert \text{ket}{0} \rangle, \lvert \text{ket}{1} \rangle, \lvert \text{ket}{+} \rangle, \lvert \text{ket}{-} \rangle, \lvert \text{ket}{R} \rangle, \lvert \text{ket}{L} \rangle$ выбраны неслучайно: они имеют геометрический смысл.

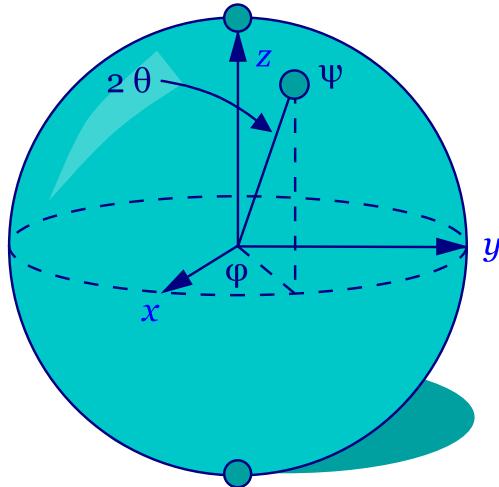


Fig. 53 Сфера Блоха

Принято считать, что ось $\lvert \text{mathbf}{Z} \rangle$ – это основная ось, так как физически квантовые компьютеры измеряют именно по ней. Ось $\lvert \text{mathbf}{X} \rangle$ “смотрит на нас” и поэтому обозначается $\lvert \text{ket}{+} \rangle$ и $\lvert \text{ket}{-} \rangle$. А ось $\lvert \text{mathbf}{Y} \rangle$ направлена как бы вдоль, поэтому базис обозначают как “право” ($\lvert \text{ket}{R} \rangle$) и “лево” ($\lvert \text{ket}{L} \rangle$).

Вектор состояния кубита еще называют волновой функцией, и этот вектор может идти в любую точку сферы Блоха. Сама сфера имеет единичный радиус, и это гарантирует нам, что для всех состояний сумма квадратов амплитуд будет равна единице.

Состояние в полярных координатах

Состояние кубита можно выразить через полярные координаты на сфере Блоха:

$$\lvert \text{ket}{\Psi} \rangle = c_0 \lvert \text{ket}{0} \rangle + c_1 \lvert \text{ket}{1} \rangle = \cos\theta \lvert \text{ket}{0} \rangle + e^{i\phi} \sin\theta \lvert \text{ket}{1} \rangle$$

где (θ, ϕ) – это угловые координаты на сфере Блоха. В этом смысле сфера Блоха очень удобна для представления состояний одного кубита.

Note

Тут мы воспользовались формулой Эйлера, а также вынесли за скобки локальные фазы множителей c_0 и c_1 . Если у вас возникают трудности с подобными операциями над комплексными числами, то рекомендуем еще раз пересмотреть базовый [блок](#) нашего курса по линейной алгебре и комплексным числам, там эти моменты освещаются более подробно.

Что можно делать с таким кубитом?

Линейные операторы

Любое действие, которое мы совершаем с кубитом в состоянии $|\Psi\rangle$, должно переводить его в другое состояние $|\Phi\rangle$. Что переводит один вектор в другой вектор в том же пространстве? Правильно, матрица. Другими словами, линейный оператор. Мы будем обозначать операторы как \hat{U} .

Унитарность

Как мы уже говорили, квадраты амплитуд – это вероятности. Следовательно, волновая функция должна быть нормирована на единицу. А значит, любой оператор, который переводит одно состояние в другое $|\hat{U}\rangle|\Psi\rangle = |\Phi\rangle$, должен сохранять эту нормировку, то есть должен быть [унитарным](#). Более того, свойство унитарности приводит к тому, что любой квантовый оператор еще и сохраняет скалярное произведение: $|\langle \Psi | \hat{U}^\dagger | \Psi \rangle| = |\langle \Psi | \hat{U} | \Psi \rangle|$

Другими словами, унитарный оператор удовлетворяет условию $|\hat{U}^\dagger| = |\hat{U}|$.

Обратимость

Одно из важных следствий унитарности операций над кубитами – это их обратимость. Если вы сделали какую-то последовательность унитарных операций над кубитами $|\hat{U}_1, \dots, \hat{U}_n\rangle$, то их можно вернуть в начальное состояние, ведь у унитарного оператора всегда есть обратный оператор $|\hat{U}_1^{-1}, \dots, \hat{U}_n^{-1}\rangle = |\hat{U}_n^\dagger, \dots, \hat{U}_1^\dagger\rangle$.

Note

Квантовый компьютер должен уметь делать несколько не унитарных операций, например, инициализацию кубита в определенное состояние (например, $|0\rangle$) и считывание состояния кубитов. Такие неунитарные операции приводят к потере информации и являются необратимыми.

Пример оператора

В дальнейших лекциях мы разберем много операторов, так как именно операторы (или квантовые **гейты**) являются основой квантовых вычислений. А пока рассмотрим простой пример: оператор Адамара (**Hadamard gate**), который переводит $|0\rangle$ в $|+\rangle$.

Гейт Адамара

Начнем с того, что пока у нас лишь один кубит. Состояние одного кубита – это вектор размерности два. Значит, оператор, который переводит его в другой вектор размерности два – это матрица 2×2 . Запишем оператор Адамара в матричном виде, а потом убедимся, что он унитарный и действительно переводит состояние $|0\rangle$ в $|+\rangle$.

```
\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}
```

Реализация в Python

```
h = 1 / np.sqrt(2) * np.array([
    [1 + 0j, 1 + 0j],
    [1 + 0j, 0j - 1]
])
```

Унитарность

```
print(np.allclose(np.conj(h).T @ h, np.eye(2)))
```

True

Проверка

Проверим, что гейт Адамара действительно переводит кубит из состояния $|\langle\text{ket}\{0\}\rangle$ в состояние $|\langle\text{ket}\{+\}\rangle$.

```
print(np.allclose(h @ basis_0, plus))
```

True

Измерение

Измерение в квантовых вычислениях выделяется отдельно именно потому, что оно “открывает” коробку с котом Шредингера: мы точно узнаем, жив он или мертв, и уже никогда не сможем это “забыть” обратно. Вся суперпозиция его состояния исчезает. То есть **измерение** – это как раз пример одной из не унитарных операций, которые должен уметь делать квантовый компьютер.

Note

Это интересный факт: исчезновение суперпозиции многим кажется парадоксом, именно поэтому и появляются разные интерпретации квантовой механики, например, многомировая интерпретация Эверетта. Действительно, это кажется немного странным, что полностью обратимая квантовая механика и непрерывная динамика волновых функций вдруг “ломаются” и мы получаем такой коллапс, который еще называют редукцией фон Неймана. Доктору Эверетту тоже это не нравилось и он предложил другую интерпретацию этого процесса. Согласно его теории, когда мы производим измерения, мы как бы “расщепляем” нашу вселенную на две ниточки: в одной кот остается жив, а в другой остается мертв.

Такие теории остаются на уровне спекуляций, так как почти невозможно придумать эксперимент, который бы подтверждал или опровергал такую гипотезу. Скорее это вопрос личного понимания и интерпретации процесса, так как математически подобные теории в итоге дают один и тот же наблюдаемый и измеримый результат.

Как мы уже говорили, состояние кубита может быть записано в разных базисах: $|\langle\text{ket}\{0\}, \text{ket}\{1\}\rangle$, $|\langle\text{ket}\{+\}, \text{ket}\{-\}\rangle$, $|\langle\text{ket}\{R\}, \text{ket}\{L\}\rangle$. Значение кубита в каждом из этих базисов может быть измерено. Но что такое измерение с точки зрения математики?

Операторы Паули

На самом деле, любая наблюдаемая величина соответствует какому-то оператору. Например, измерения в разных базисах $|\langle\text{mathbf}\{X\}\rangle$, $|\langle\text{mathbf}\{Y\}\rangle$, $|\langle\text{mathbf}\{Z\}\rangle$ соответствуют операторам Паули:

$$\begin{aligned} \hat{\sigma}_x &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ \hat{\sigma}_y &= \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \\ \hat{\sigma}_z &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \end{aligned}$$

```
 pauli_x = np.array([[0 + 0j, 1 + 0j], [1 + 0j, 0 + 0j]])
 pauli_y = np.array([[0 + 0j, 0 - 1j], [0 + 1j, 0 + 0j]])
 pauli_z = np.array([[1 + 0j, 0 + 0j], [0 + 0j, 0j - 1j]])
```

Эти операторы очень важны, рекомендуется знать их наизусть, так как они встречаются в каждой второй статье по квантовым вычислениям, а также постоянно фигурируют в документации всех основных библиотек для квантового машинного обучения.

Собственные значения

Мы поняли, что есть связь между нашими измерениями и операторами. Но какая именно? Что значит, например, что измерения по оси $|\langle\text{mathbf}\{Z\}\rangle$ соответствуют оператору $|\langle\hat{\sigma}_z\rangle|$?

Здесь мы приходим к собственным значениям операторов. Оказывается (так устроен наш мир), что *измеряя* какую-то величину в квантовой механике, мы всегда будем получать одно из собственных значений соответствующего оператора, а состояние будет колапсировать в соответствующий собственный вектор этого оператора. Другими словами, *измеряя* кота Шредингера, мы будем получать значения “жив” или “мертв”, а

состоение кота будет переходить в состояние, соответствующее одному из этих значений. А еще измерение не является обратимой операцией: однажды открыв коробку с котом и поняв, жив он или мертв, мы уже не сможем закрыть ее обратно и вернуть кота в суперпозицию.

Описанное выше – не абстрактные рассуждения из квантовой физики. Оно пригодится, когда мы будем говорить о решении практических комбинаторных задач, таких как задача о выделении сообществ в графе.

Собственные вектора $\langle \hat{\sigma}^z \rangle$

Вернемся к нашему оператору $\langle \hat{\sigma}^z \rangle$. Легко убедиться, что его собственные значения равны 1 и -1, а соответствующие им собственные вектора – это $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ и $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$:

```
print(np.linalg.eig(pauli_z))

(array([ 1.+0.j, -1.+0.j]), array([[1.+0.j, 0.+0.j],
[0.+0.j, 1.+0.j]]))
```

Таким образом, измерение по оси $\langle \mathbf{Z} \rangle$ всегда будет давать нам одно из этих двух значений и переводить состояние кубита в соответствующий собственный вектор.

⚠ Caution

Зачастую кубиты измеряют именно в $\langle \mathbf{Z} \rangle$ -базисе; является неким "стандартом" для квантовых вычислений, так как это измерение "ближе к железу". Также $\langle \mathbf{Z} \rangle$ -базис удобен для нас из-за диагональности [оператора Паули](#) $\langle \sigma^z \rangle$.

Формальная запись

Формально мы можем записать для любого [эрмитова](#) оператора $\langle \hat{U} \rangle$, что собственные состояния этого оператора являются его собственными векторами, а собственные значения в этом случае являются наблюдаемыми значениями:

$$\langle \hat{U} \rangle \langle \Psi \rangle = u \langle \Psi \rangle$$

Другие операторы Паули

Убедимся, что у остальных операторов собственные значения такие же:

```
print(np.linalg.eig(pauli_x))
print(np.linalg.eig(pauli_y))

(array([ 1.+0.j, -1.+0.j]), array([[ 0.70710678-0.j,  0.70710678+0.j],
[ 0.70710678+0.j, -0.70710678-0.j]]))
(array([ 1.+0.j, -1.+0.j]), array([[[-0.          , -0.70710678j,  0.70710678+0.j],
[ 0.70710678+0.j,  0.          , -0.70710678j]]]))
```

Note

Заметим, что собственные вектора могут отличаться на какой-то множитель. В частности, один из собственных векторов оператора $\hat{\sigma}_y$, возвращенный `np.linalg.eig` равен $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}$, что отличается от $\ket{R} = \frac{1}{\sqrt{2}} (\ket{0} + i\ket{1}) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \end{bmatrix}$ только домножением на i .

Во-первых, по курсу линейной алгебры мы знаем, что собственные вектора можно умножать на любую константу и они все равно останутся собственными, поскольку если вектор (x^*) – решение уравнения $(Ax^* = \lambda x^*)$, то и любой вектор $(kx^*, k \in \mathbb{C}, k \neq 0)$ также будет решением этого уравнения.

Во-вторых, состояния, отличающиеся только множителем, по сути соответствуют одному и тому же состоянию. Любой множитель $(c \in \mathbb{C})$, стоящий перед вектором состояния $(\ket{\Psi})$ можно представить в виде $(c = e^{i\phi})$, что соответствует повороту на некоторый угол (ϕ) . При вычислении амплитуды $(|\langle \Psi | \Psi \rangle|^2)$ этот множитель даст единицу, то есть, другими словами, поворот кубита никак не влияет на результат измерения кубита. Это известно как *глобальная фаза*, и в научной литературе часто можно встретить фразу “up to a global phase factor”, что означает одинаковые состояния с точностью до поворота (примерно как произвольная константа, добавляемая к интегралу).

Note

Также можно заметить, что у всех операторов Паули нет ни одного общего собственного вектора. Таким образом, мы приходим к ситуации, когда не можем одновременно точно провести измерения двумя разными операторами, так как наше измерение должно переводить состояние в соответствующий собственный вектор. В квантовой механике это называется **принципом неопределенности**.

Ожидаемое значение при измерении

Мы не будем писать с нуля полный симулятор кубитов, который включает измерения – это требует введения сложного случайного процесса. Но мы можем легко ответить на другой вопрос. А именно: можно ли сказать, какое будет ожидаемое значение оператора (\hat{U}) для состояния $(|\Psi\rangle)$? Другими словами, какое будет математическое ожидание большого числа измерений? Это можно записать следующим образом:

$$\mathbf{E}(\hat{U}) = \bra{\Psi} \hat{U} \ket{\Psi}$$

Например, оператор $(\hat{\sigma}_z)$ полностью не определен в состоянии $(\ket{+})$, то есть мы будем равновероятно получать значения -1 и 1 , а математическое ожидание, соответственно, будет равно нулю:

```
print(plus.conj().T @ pauli_z @ plus)
```

```
[[ -2.23711432e-17+0.j]]
```

С другой стороны, измеряя состояние $(\ket{+})$ в X -базисе мы всегда будем получать 1 :

```
print(plus.conj().T @ pauli_x @ plus)
```

```
[[1.+0.j]]
```

Вероятности битовых строк

Последнее, чего мы коснемся в части измерений – это битовые строки и метод Шредингера. Мы много говорили о вероятностной интерпретации волновой функции и аналогиях с классическим битом, но пока этого никак не касались на практике. Как же получить вероятность определенной битовой строки для произвольного состояния? Если взять все битовые строки размерности вектора состояния и отсортировать их в лексикографическом порядке (например, $(0 < 1)$, $(00 < 01 < 10 < 11)$, и т.д.), то вероятность каждой битовой строки получается следующим выражением:

$\mathbf{P} = |\bra{\Psi} \vec{v} \ket{\Psi}|^2$

где (\vec{v}) – это вектор, каждая компонента которого соответствует порядковой битовой строке или вектор битовых строк. Другими словами, вероятность получить i -ю битовую строку равна квадрату i -го элемента амплитуды волновой функции. Кажется немного запутанным, но на самом деле $(|\langle \Psi | \Psi \rangle|^2)$ – это идеально и есть плотность вероятности.

Еще пара слов об измерениях

Измерение как проекция на пространство собственных векторов

Мы уже говорили, что при измерении мы как бы "выбираем" один из собственных векторов наблюдаемой. Более строго такой процесс называется проецированием на пространство собственных векторов. Для собственного вектора $\langle \hat{P} | \Phi_i \rangle$ проекция будет линейным оператором:

$$\langle \hat{P} | \Phi_i \rangle = \langle \Phi_i | \hat{P} | \Phi_i \rangle$$

```
super_position = h @ basis_0
eigenvectors = np.linalg.eig(pauli_z)[1]

proj_0 = eigenvectors[0].reshape((-1, 1)) @ eigenvectors[0].reshape((1, -1))
proj_1 = eigenvectors[1].reshape((-1, 1)) @ eigenvectors[1].reshape((1, -1))
```

Правило Борна

Как мы уже говорили, любому эрмитову оператору соответствует какая-либо наблюдаемая величина. А какая наблюдаемая величина соответствует оператору проекции на собственный вектор $\langle \Phi_i | \hat{P} | \Phi_i \rangle$, про который мы говорили выше? Ответ – вероятность наблюдения собственного значения, которое соответствует этому собственному вектору. Значит, чтобы получить вероятность измерения значения λ_i эрмитова оператора \hat{P} (которое соответствует собственному вектору $\langle \Phi_i | \hat{P} | \Phi_i \rangle$ этого оператора) в состоянии $|\Psi\rangle$, мы должны измерить величину $|\langle \Phi_i | \hat{P} | \Psi \rangle|^2$. Это называется правилом Борна.

$$|\langle \Phi_i | \hat{P} | \Psi \rangle|^2 = |\langle \Phi_i | \Psi \rangle|^2$$

Считать ожидаемое значение оператора мы уже умеем. Давайте убедимся, что для состояния $|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ результаты измерений операторов проекций дадут 0.5 и совпадут с результатом упражнения, которое мы проделали ранее:

```
p_0 = super_position.conj().T @ proj_0 @ super_position
p_1 = super_position.conj().T @ proj_1 @ super_position

print(np.allclose(p_0 + p_1, 1.0))
print(np.allclose(p_0, 0.5))
```

```
True
True
```

Что мы узнали?

- Состояние и значение для кубита – это не одно и то же.
- Состояния представляют собой комплекснозначные вектора.
- Квантовые операторы – унитарные и самосопряженные.
- Измеряемые значения – собственные значения операторов.
- Измерение "ломает" суперпозицию.

Квантовые гейты

Описание лекции

Из этой лекции мы узнаем:

- какие есть основные однокубитные и многокубитные гейты;
- как записывать многокубитные состояния;
- как конструировать многокубитные операторы;
- как работать с библиотекой [PennyLane](#).

Введение

Квантовые гейты являются основными строительными блоками для любых квантовых схем, в том числе и тех, что применяются для машинного обучения. Можно сказать, что это своеобразный алфавит квантовых вычислений. Он необходим, чтобы сходить понимать, например, что изображено на подобных схемах:



FIG. 2: L-VQE ansatz for a 6-qubit quantum state. R_y denotes rotation around the y -axis defined as $R_y(\theta) \equiv e^{-i\frac{\theta}{2}Y}$. Every R_y contains a parameter that is optimized over in the outer loop.

Fig. 54 Схема Layered-VQE

Основные однокубитные гейты

В прошлый раз мы познакомились с [операторами Паули](#), а также [гейтом Адамара](#). Как для обычных квантовых алгоритмов, так и для QML-алгоритмов нужны и другие гейты, потому что одни только эти гейты не позволяют перейти во все возможные квантовые состояния. Теперь давайте посмотрим, какие еще однокубитные гейты часто применяются в квантовых вычислениях и квантовом машинном обучении.

T-гейт

T-гейт очень популярен в универсальных квантовых вычислениях. Его матрица имеет вид:

$$\begin{bmatrix} \hat{T} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1+i}{\sqrt{2}} \end{bmatrix} \end{bmatrix}$$

Любой однокубитный гейт можно аппроксимировать последовательностью гейтов Адамара и T-гейтов. Чем точнее требуется аппроксимация, тем длиннее будет аппроксимирующая последовательность.

Помимо важной роли в математике квантовых вычислений, гейт Адамара и T-гейт интересны тем, что именно на них построено большинство предложений по реализации квантовых вычислений с топологической защитой или с коррекцией ошибок. На сегодняшний день эти схемы реально пока не очень работают: никаких топологически защищенных кубитов продемонстрировано не было, а коррекция ошибок не выходит за пределы двух логических кубитов.

Гейты поворота вокруг оси

Поворотные гейты играют центральную роль в квантовом машинном обучении. Вспомним на секунду, как выглядят наши однокубитные состояния на сфере Блоха:

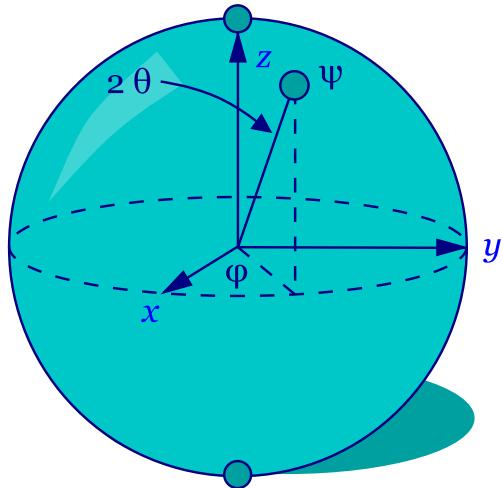


Fig. 55 Сфера Блоха

Любой однокубитный гейт можно представить как вращение вектора состояния $|\Psi\rangle$ на некоторый угол вокруг некоторой оси, проходящей через центр сферы Блоха.

Гейты $\hat{R}X(\phi)$, $\hat{R}Y(\phi)$, $\hat{R}Z(\phi)$ осуществляют поворот на определенный угол ϕ вокруг соответствующей оси на сфере Блоха.

Давайте внимательно рассмотрим это на примере гейта $\hat{R}Y(\phi)$.

Гейт $\hat{R}Y(\phi)$

Сам гейт определяется следующим образом:

$$\begin{aligned} \hat{R}Y(\phi) = & \begin{pmatrix} \cos(\frac{\phi}{2}) & -\sin(\frac{\phi}{2}) \\ \sin(\frac{\phi}{2}) & \cos(\frac{\phi}{2}) \end{pmatrix} \end{aligned}$$

```
import numpy as np

def ry(state, phi):
    return np.array([
        [np.cos(phi / 2), -np.sin(phi / 2)],
        [np.sin(phi / 2), np.cos(phi / 2)]
    ]) @ state
```

Запишем наше состояние $|0\rangle$:

```
basis = np.array([1 + 0j, 0 + 0j]).reshape((2, 1))
```

Внимательно посмотрим на сферу Блоха. Можно заметить, что если повернуть состояние из $|0\rangle$ на $\frac{\pi}{2}$ и измерить значение $\langle \sigma_x \rangle$, то получится 1. А если повернуть на $-\frac{\pi}{2}$, то получится -1:

```
def expval(state, op):
    return state.conj().T @ op @ state

pauli_x = np.array([[0 + 0j, 1 + 0j], [1 + 0j, 0 + 0j]])

print(np.allclose(expval(ry(basis, np.pi / 2), pauli_x), 1.0))
print(np.allclose(expval(ry(basis, -np.pi / 2), pauli_x), -1.0))
```

```
True
True
```

Убедимся также, что вращение на угол, пропорциональный 2π , не меняет результат измерения. Возьмем случайное состояние:

$$|\Psi\rangle = \begin{pmatrix} 0.42 \\ \sqrt{1 - 0.42^2} \end{pmatrix}$$

```
random_state = np.array([0.42 + 0j, np.sqrt(1 - 0.42**2) + 0j]).reshape((2, 1))
```

Измерим его по осям X и Z , затем повернем на угол 2π и измерим снова:

```
pauli_z = np.array([[1 + 0j, 0 + 0j], [0 + 0j, 0j - 1]])

print("Z:\n" + str(expval(random_state, pauli_z)) + "\n")
print("X:\n" + str(expval(random_state, pauli_x)) + "\n")

print("Z after RY:\n" + str(expval(ry(random_state, 2 * np.pi), pauli_z)) + "\n")
print("X after RY:\n" + str(expval(ry(random_state, 2 * np.pi), pauli_x)) + "\n")
```

```
Z:
 [[-0.6472+0.j]]

X:
 [[0.76232025+0.j]]

Z after RY:
 [[-0.6472+0.j]]

X after RY:
 [[0.76232025+0.j]]
```

Другие гейты вращений

Аналогичным образом определяются гейты $\hat{R}X(\phi)$ и $\hat{R}Z(\phi)$:

$$\begin{aligned} \hat{R}X(\phi) = & \begin{pmatrix} \cos(\frac{\phi}{2}) & -i\sin(\frac{\phi}{2}) \\ i\sin(\frac{\phi}{2}) & \cos(\frac{\phi}{2}) \end{pmatrix} \\ \hat{R}Z(\phi) = & \begin{pmatrix} e^{-i\frac{\phi}{2}} & 0 \\ 0 & e^{i\frac{\phi}{2}} \end{pmatrix} \end{aligned}$$

Общая форма записи однокубитных гейтов

В общем случае однокубитные гейты могут быть также записаны следующим образом:

$$\begin{aligned} \left[\begin{array}{c} \hat{R}^{\alpha} \\ \vec{n} \end{array} \right] &= e^{-i\frac{\alpha}{2}\hat{\sigma}_z} \left[\begin{array}{c} \vec{n} \\ \hat{\sigma}_z \end{array} \right] e^{i\frac{\alpha}{2}\hat{\sigma}_z} \\ \text{где } \hat{\alpha} &- \text{это угол поворота, } \vec{n} - \text{единичный вектор в направлении оси поворота, а } \hat{\sigma}_z \\ &= \hat{\sigma}_z^x, \hat{\sigma}_z^y, \hat{\sigma}_z^z - \text{это вектор, составленный из операторов Паули. Если} \\ &\text{использовать покоординатную запись и } \vec{n} = [n_x, n_y, n_z] \text{ задает ось вращения, то} \\ \left[\begin{array}{c} \hat{R}^{\alpha} \\ \vec{n} \end{array} \right] &= e^{-i\frac{\alpha}{2}} \left(\hat{\sigma}_z^x n_x + \hat{\sigma}_z^y n_y + \hat{\sigma}_z^z n_z \right) \end{aligned}$$

Забегая вперед, можно сказать, что именно гейты вращений – это основа [квантовых вариационных схем](#), главного инструмента этого курса.

Phase-shift гейт

Другой важный гейт – это так называемый phase-shift гейт, или \hat{U}_1 гейт. Его матричная форма имеет следующий вид:

$$\begin{aligned} \left[\begin{array}{c} \hat{U}_1(\phi) \end{array} \right] &= \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix} \end{aligned}$$

```
def u1(state, phi):
    return np.array([[1, 0], [0, np.exp(1j * phi)]]) @ state
```

Легко видеть, что с точностью до глобального фазового множителя, который ни на что не влияет, Phase-shift-гейт – это тот же $\hat{R}_Z(\phi)$. Он играет важную роль в квантовых ядерных методах.

Гейты \hat{U}_2 и \hat{U}_3

Более редкие в QML гейты, которые однако все равно встречаются в статьях.

$$\begin{aligned} \left[\begin{array}{c} \hat{U}_2(\phi, \lambda) \end{array} \right] &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -e^{i\lambda} e^{i\phi} \\ e^{i\lambda} e^{i\phi} & 1 \end{bmatrix} \end{aligned}$$

Давайте убедимся в справедливости этого выражения:

```
def rz(state, phi):
    return np.array([[np.exp(-1j * phi / 2), 0], [0, np.exp(1j * phi / 2)]]) @ state

def u2_direct(phi, l):
    return (
        1
        / np.sqrt(2)
        * np.array([[1, -np.exp(1j * l)], [np.exp(1j * phi), np.exp(1j * (phi + l))]])
    )

def u2_inferenced(phi, l):
    return (
        u1(np.eye(2), phi + l)
        @ rz(np.eye(2), -l)
        @ ry(np.eye(2), np.pi / 2)
        @ rz(np.eye(2), l)
    )

print(np.allclose(u2_direct(np.pi / 6, np.pi / 3), u2_inferenced(np.pi / 6, np.pi / 3)))
```

True

Схожим образом определяется $\hat{U}_3(\theta, \phi, \lambda)$:

$$\begin{aligned} \left[\begin{array}{c} \hat{U}_3(\theta, \phi, \lambda) \end{array} \right] &= \begin{bmatrix} \cos(\frac{\theta}{2}) & -e^{i\phi} \sin(\frac{\theta}{2}) & 0 \\ e^{i\phi} \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) & 0 \\ 0 & 0 & e^{i\lambda} \end{bmatrix} \hat{U}_1(\phi + \lambda) \hat{R}_Z(-\lambda) \hat{R}_Y(\theta) \hat{R}_Z(\lambda) \end{aligned}$$

Читатель может сам легко убедиться, что эти формы записи эквивалентны. Для этого надо написать примерно такой же код, что мы писали раньше для \hat{U}_2 .

Еще пара слов об однокубитных гейтах

На этом мы завершаем обзор основных однокубитных гейтов. Маленько замечание: гейты, связанные со сдвигом фазы, никак не меняют состояние кубита, если оно сейчас $|0\rangle$. Так как мы всегда предполагаем, что начальное состояние кубитов – это именно $|0\rangle$, то перед применением, например, \hat{U}_1 , рекомендуется применить гейт Адамара:

```

print(np.allclose(u1(basis, np.pi / 6), basis))
h = 1 / np.sqrt(2) * np.array([[1 + 0j, 1 + 0j], [1 + 0j, 0j - 1j]])
print(np.allclose(u1(h @ basis, np.pi / 6), h @ basis))

```

```

True
False

```

Единичный гейт

Самое последнее об однокубитных гейтах – это единичный гейт \hat{I} :

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

```

identity_gate = np.eye(2, dtype=np.complex128)
print(identity_gate)

```

```

[[1.+0.j 0.+0.j]
 [0.+0.j 1.+0.j]]

```

Он не делает с кубитом ровным счетом ничего. Но единичный гейт понадобится нам позже, когда мы будем конструировать многокубитные операторы.

Многокубитные состояния и гейты

Очевидно, что с одним кубитом ничего интересного, кроме разве что генератора истинно-случайных чисел, мы не сделаем. Для начала разберемся, как выглядят состояния для многокубитных систем.

Многокубитные состояния

В классическом компьютере один бит имеет два значения – 0 и 1. Два бита имеют четыре значения – 00, 01, 10, 11. Три бита имеют восемь значений и так далее. Аналогично состояние двух кубитов – это вектор в пространстве \mathbb{C}^4 , состояние трех кубитов – вектор в пространстве \mathbb{C}^8 , то есть состояние N кубитов описывается вектором размерности 2^N в комплексном пространстве. Вероятности каждой из возможных битовых строк ($0000\dots000$, $0000\dots010$, $0000\dots100$, и так далее) получаются по методу Шредингера, который мы обсуждали [в конце прошлой лекции](#):

$$|\Psi\rangle = |\psi_0\rangle\langle\psi_0| + |\psi_1\rangle\langle\psi_1| + \dots + |\psi_{2^N-1}\rangle\langle\psi_{2^N-1}|$$

Нужно отсортировать наши битовые строки в лексикографическом порядке – и вероятность i -й битовой строки будет равна квадрату i -го элемента вектора $|\Psi\rangle$.

Формально, многокубитные состояния описываются с помощью математического концепта так называемого [тензорного произведения](#), которое в случае линейных операторов идентично [произведению Кронекера](#), обозначаемого знаком \otimes . Так, если $|\Psi_A\rangle$ и $|\Psi_B\rangle$, то $|\Psi_{AB}\rangle = |\Psi_A\rangle \otimes |\Psi_B\rangle$. О том, как элементы вектора $|\Psi_{AB}\rangle$ выражаются через элементы векторов $|\Psi_A\rangle$ и $|\Psi_B\rangle$, можно прочитать на Википедии в статье [“Произведение Кронекера”](#).

Многокубитные операторы

Как мы уже обсуждали ранее, квантовые операторы должны переводить текущее состояние в новое в том же пространстве и сохранять нормировку, а еще должны быть обратимыми. Значит, оператор для состояния из N кубитов – это унитарная комплексная матрица размерности $(2^N \times 2^N)$.

Конструирование многокубитных операторов

Прежде чем мы начнем обсуждать двухкубитные операторы, рассмотрим ситуацию. Представим, что у нас есть состояние из двух кубитов и мы хотим подействовать на первый кубит оператором Адамара. Как же тогда нам написать такой двухкубитный оператор? Мы знаем, что действуем на первый кубит оператором, а что происходит со вторым кубитом? Ничего не происходит – и это эквивалентно тому, что мы действуем на второй кубит единичным оператором. А финальный оператор $(2^2 \times 2^2)$ записывается через произведение Кронекера:

$$\begin{aligned} \hat{H} \otimes \hat{I} &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{aligned}$$

Учитывая, что многокубитные состояния конструируются аналогичным образом через произведение Кронекера, мы можем убедиться в верности нашего вывода:

```
print(np.allclose(np.kron(h @ basis, basis), np.kron(h, identity_gate) @ np.kron(basis, basis)))
```

```
True
```

Наблюдаемые для многокубитных гейтов

Аналогичным образом можно сконструировать и наблюдаемые. Например, если мы хотим измерять одновременно два спина по оси \hat{Z} , то наблюдаемая будет выглядеть так:

$$\begin{bmatrix} \hat{Z}\hat{Z} = \hat{\sigma}_z\hat{\sigma}_z & \\ \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

```
print(np.kron(basis, basis).conj().T @ np.kron(pauli_z, pauli_z) @ np.kron(basis, basis))
```

```
[[1.+0.j]]
```

Основные двухкубитные гейты

Основные многокубитные гейты, которые предоставляют современные квантовые компьютеры, – это двухкубитные гейты.

CNOT (CX)

Квантовый гейт контролируемого инвертирования – это гейт, который действует на два кубита: *рабочий* и *контрольный*. В зависимости от того, имеет ли контрольный кубит значение 1 или 0, этот гейт инвертирует или не инвертирует рабочий кубит.

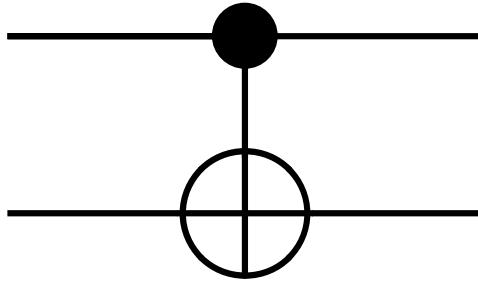


Fig. 56 Гейт CNOT

Иногда этот гейт также называют гейтом CX. В матричном виде этот оператор можно записать так:

$$\hat{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

```
cnot = (1 + 0j) * np.array([
    [1, 0, 0, 0],
    [0, 1, 0, 0],
    [0, 0, 0, 1],
    [0, 0, 1, 0],
])
print(np.allclose(cnot @ np.kron(basis, basis), np.kron(basis, basis)))
print(np.allclose(
    cnot @ np.kron(pauli_x @ basis, basis), np.kron(pauli_x @ basis, pauli_x @ basis)
))
```

```
True  
True
```

Заметьте, тут мы воспользовались тем, что $\hat{\sigma}_x$ работает так же, как инвертор кубитов: он превращает $|0\rangle$ в $|1\rangle$ и наоборот.

Гейты CY и CZ

Схожие по принципу гейты – это гейты \hat{CY} и \hat{CZ} . В зависимости от значения управляющего кубита к рабочему кубиту применяют соответствующий оператор Паули:

$$\begin{aligned} \hat{CY} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{bmatrix} \\ \hat{CZ} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \end{aligned}$$

Гейт iSWAP

Гейты \hat{CX} , \hat{CY} и \hat{CZ} эквивалентны с точностью до однокубитных гейтов. Это означает, что любой из них можно получить, добавив необходимые однокубитные гейты до и после другого гейта. Например:

$$[\hat{CZ}] = [\hat{H} \hat{X} \hat{H}]$$

Этим свойством обладают отнюдь не все двухкубитные гейты. Например, таковым является гейт iSWAP:

$$[\hat{iSWAP}] = [\hat{H} \hat{X} \hat{H} \hat{X} \hat{H}]$$

Гейт fSim

Для разных архитектур квантовых процессоров "естественный" гейт может выглядеть по-разному. Например, в квантовом процессоре Google Sycamore естественным является так называемый fermionic simulation gate или fSim. Это двухпараметрическое семейство гейтов вида:

$$[\hat{fSim}(\theta, \phi)] = [\cos(\theta) \cos(\phi) \sin(\theta) \sin(\phi)]$$

Впрочем, и fSim-гейт не является эквивалентным всему множеству двухкубитных гейтов. В общем случае, чем больше кубитов, тем сложнее будет выглядеть декомпозиция произвольного гейта на физически реализуемые "железе".

Первое знакомство с PennyLane

На сегодняшний день существует достаточно много фреймворков для программирования квантовых компьютеров. Для целей этого курса мы будем использовать [PennyLane](#). Эта библиотека предоставляет высокогенеральный [Python API](#) и создана специально для решения задач квантового машинного обучения.

```
import pennylane as qml
```

Device

Для объявления квантового устройства используется класс `Device`. PennyLane поддерживает работу с большинством существующих квантовых компьютеров, но для целей курса мы будем запускать все наши программы лишь на самом простом симуляторе идеального квантового компьютера:

Note

Разработка и отладка квантовых алгоритмов, как правило, происходит на симуляторах. Но надо понимать, что это работает только пока алгоритмы – "игрушечные" и задействуют пару-тройку или пару десятков кубитов. Надо понимать, что при добавлении каждого следующего кубита требуется вдвое больше ресурсов, чтобы просимулировать квантовый компьютер. Поэтому симуляторы – принципиально плохо масштабируемые аналоги реальных квантовых компьютеров.

```
device = qml.device("default.qubit", 2)
```

Первый аргумент тут – указание устройства, а второй – число кубитов.

QNode

Основной строительный блок в PennyLane – это `qnode`. Это функция, которая отмечена специальным декоратором и включает в себя несколько операций с кубитами. Результатом такой функции всегда является измерение. Напишем функцию, которая поворачивает первый кубит на (45°) , после чего измеряет оба кубита по оси \mathbf{Z} .

Сначала на NumPy

```
state = np.kron(basis, basis)
op = np.kron(ry(np.eye(2), np.deg2rad(45)), np.eye(2, dtype=np.complex128))
measure = np.kron(pauli_z, pauli_z)

print((op @ state).conj().T @ measure @ (op @ state))
```

```
[[0.70710678+0.j]]
```

Теперь через QNode

```
@qml.qnode(device)
def test(angle):
    qml.RY(angle, wires=0)
    return qml.expval(qml.PauliZ(0) @ qml.PauliZ(1))

print(test(np.deg2rad(45)))
```

```
0.7071067811865475
```

Заключение

Это последняя вводная лекция, где мы сами писали операторы и операции на чистом NumPy: это должно помочь лучше понять ту математику, которая лежит “под капотом” у квантовых вычислений. Дальше мы будем пользоваться только PennyLane и в отдельной лекции расскажем, как работать с этим фреймворком.

Итого:

- мы знаем, что такое кубит;
- понимаем линейную алгебру, которая описывает квантовые вычисления;
- понимаем, как можно сконструировать нужный нам оператор и как его применить;
- знаем, что такое измерение и наблюдаемые.

Теперь мы готовы к тому, чтобы знакомиться с квантовыми вариационными схемами и переходить непосредственно к построению моделей квантового машинного обучения.

Задачи

- Как связаны оси и угол вращения на сфере Блоха с собственными значениями и собственными векторами матрицы однокубитного гейта? Для этого найдите собственные векторы и собственные значения гейта $R^n(\alpha)$.
- Вокруг какой оси и на какой угол вращает состояние гейт Адамара?
- Гейт SWAP меняет кубиты местами. Его унитарная матрица имеет вид:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Попробуйте составить последовательность гейтов, реализующую SWAP , из гейтов $i\text{SWAP}$, $\hat{\text{CZ}}$ и $\hat{\text{RZ}}(\phi)$.

Смешанные состояния и матрицы плотности

Описание лекции

Из этой лекции мы узнаем:

- Что такое матрица плотности
- Как ввести в описание квантовых состояний новый уровень случайности
- Как связаны смешанные (mixed) и запутанные (entangled) состояния
- Как можно описать насколько "сильно" квантовое состояние является смешанным

Смешанные состояния

Эта глава кратко рассматривает довольно сложный, но крайне важный аспект квантового описания мира: как описать квантовую систему, в которой случайным является не только результат измерения, но и само состояние квантовой системы, иными словами, когда мы не можем описать квантовую систему определенным вектором состояния.

Как могут появиться такие ситуации?

Прежде всего, в любой реалистичной ситуации квантовая система будет испытывать влияние окружающей среды. Это воздействие окружающей среды на квантовый компьютер, как правило, может быть охарактеризовано некоторой температурой этой самой среды и скоростью термализации системы с ней. В отличие от классических компьютеров, где шумы редко приводят к ошибкам в вычислениях, квантовые компьютеры очень плохо защищены от шумов. Любое масштабное квантовое вычисление является гонкой с процессами декогеренции и релаксации: за время выполнения алгоритма шумы не должны испортить результат вычислений настолько, чтобы его нельзя было использовать.

Вторая, более важная причина использования смешанных состояний состоит в их тесной связи с запутанными состояниями, как показано ниже.

Почему нельзя обойтись волновой функцией

Рассмотрим реальный физический кубит в состоянии равновесия, например, спин ядра атома в магнитном поле. Статистическая физика говорит нам, что вероятность обнаружить этот спин ориентированным вдоль внешнего поля выше, чем в противоположном полю направлении. Для расчета отношений этих вероятностей можно использовать распределение Больцмана: $\langle p_{\uparrow} | \psi \rangle = \begin{pmatrix} \sqrt{p_{\uparrow}} \\ \sqrt{p_{\downarrow}} \end{pmatrix} e^{\frac{E}{k_B T}}$, где $(E - k_B T)$ – разность энергий состояний со спином вдоль поля и в противоположном полю направлений, k_B – постоянная Больцмана, а T – температура. Так как состояний у этой системы всего два, сумма их вероятностей должна давать единицу: $|p_{\uparrow}| + |p_{\downarrow}| = 1$. Таким образом, вероятности состояний однозначно определены.

Волновую функцию, отвечающую такому тепловому состоянию, можно записать в виде

$$\begin{aligned} |\psi\rangle = \begin{pmatrix} \sqrt{p_{\uparrow}} \\ \sqrt{p_{\downarrow}} \end{pmatrix} e^{i\phi} \end{aligned}$$

 где значение фазы ϕ не определено. Однако ϕ определяет поведение системы не в меньшей мере, чем вероятности $|p_{\uparrow}|$ и $|p_{\downarrow}|$. ϕ может равновероятно принимать любые значения. Таким образом, уже даже состояние теплового равновесия нельзя описать одной волновой функцией – это будет распределением вероятности по волновым функциям с разным ϕ .

Матрица плотности

Оказывается, что вместо распределения вероятностей по волновым функциям можно использовать более простую конструкцию – матрицу плотности. Если система находится в состояниях $|\Phi_n\rangle$ с вероятностями p_n , то матрицу плотности можно определить как

$$\rho = \sum_n p_n |\Phi_n\rangle \langle \Phi_n| \quad (1)$$

Выражение $|\langle \Phi_n | \Phi_n \rangle|$ обозначает произведение вектора-столбца на вектор-строку – результатом будет матрица. Важно, что значение любой ожидаемой величины (отвечающей оператору \hat{A}) можно записать через ρ :

$$\langle \hat{A} \rangle = \sum_n p_n \langle \Phi_n | \hat{A} | \Phi_n \rangle = \sum_n p_n \operatorname{Tr}[\hat{A}]$$

Математическое обоснование этой циклической перестановки можно получить, расписав матричные произведения покомпонентно:

$$\langle \hat{A} \rangle = \sum_n p_n \sum_{ij} \langle \Phi_n | A_{ij} | \Phi_n \rangle = \operatorname{Tr}[\hat{A}]$$

Любую наблюдаемую физическую величину можно выразить в виде ожидаемой величины некоторого эрмитова оператора – а значит, описание с помощью матрицы плотности является универсальным для любых случайных квантовых систем.

Стоит заметить, что матрица плотности для подсистем была впервые введена в научный оборот знаменитым советским физиком, лауреатом Нобелевской премии Львом Ландау. [15a]

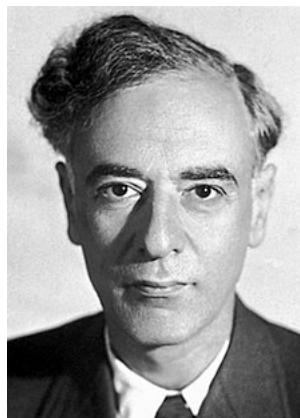


Fig. 57 Лев Ландау, 1908-1968

Чистые и смешанные состояния

Состояния, которые описываются одной единственной волновой функцией $|\Psi\rangle$, называются *чистыми* (англ. *pure states*). Для таких состояний выражение для матрицы плотности получается тривиальным:

$$|\rho = |\Psi\rangle \langle \Psi| \quad (2)$$

Матрица плотности чистого состояния является *оператором-проектором*: действие оператора на волновую функцию произвольного состояния $|\Phi\rangle$ дает проекцию $|\Phi\rangle$ на $|\Psi\rangle$. Состояния, которые нельзя описать одним вектором состояния, а можно лишь матрицей плотности, называются *смешанными* (англ. *mixed states*).

Чистота состояния

Можно легко показать, для оператора-проектора (2) выполняется тождество

$$|\rho^2 = |\Psi\rangle \langle \Psi| |\Psi\rangle \langle \Psi| = |\Psi\rangle \langle \Psi| = |\rho|$$

и следовательно, для чистого состояния

$$|\operatorname{Tr}(\rho^2) = 1|$$

(напомним, что мы всегда считаем вектора состояния нормированными на единицу).

Аналогичным образом, но после более длинных выкладок можно показать, что в общем случае

$$|\operatorname{Tr}(\rho^2) \leq 1|$$

причем знак равенства в последней формуле возможен, только если в формуле (1) сумма имеет лишь одно слагаемое (т.е. состояние является чистым). Последнее свойство матрицы плотности позволяет ввести ряд величин, характеризующих смешанные и запутанные состояния, о чем будет рассказано в заключительном разделе этой лекции. Величина

$$|\gamma = \operatorname{Tr}(\rho^2)| \quad (3)$$

называется *чистотой состояния* (*quantum state purity*).

Энтропия

Энтропия фон Неймана – это другая численная характеристика того, насколько сильно наша система смешанная.

Ее выражение очень похоже на выражение для классической [энтропии Шеннона](#). Только в отличие от классики, в квантовой механике мы имеем матрицу плотности, поэтому в выражении у нас фигурирует [матричный логарифм](#):

$$|S = -\operatorname{Tr}(\rho \ln(\rho))| \quad (4)$$

Спектральная декомпозиция матрицы плотности

Определение матрицы плотности (1) представляет собой сумму матриц плотностей чистых состояний, взятых с некоторыми вероятностями. Интересно, что совершенно разным комбинациям чистых состояний могут соответствовать одинаковые матрицы плотности. Например, состояние кубита, который с вероятностью 50% находится в состоянии $|\Psi_0\rangle$ и с вероятностью 50% в состоянии $|\Psi_1\rangle$ совершенно неотличимо от такой же равновероятной смеси состояний $|\Psi_+\rangle$ и $|\Psi_-\rangle$:

$$\begin{aligned} & \left| \frac{1}{2}(|\Psi_0\rangle \langle \Psi_0| + |\Psi_1\rangle \langle \Psi_1|) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \right| \\ & \left| \frac{1}{2}(|\Psi_+\rangle \langle \Psi_+| + |\Psi_-\rangle \langle \Psi_-|) = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \right| \end{aligned}$$

Соотношение (1) также можно рассматривать как спектральную декомпозицию матрицы плотности. В этом случае состояния (Ψ_i) образуют ортонормированный базис, а вероятности (p_i) – это собственные значения ρ . Как чистота (3), так и энтропия (4) зависят лишь от этих собственных значений.

Смешанные состояния и запутанность

Рассмотрим ситуацию, когда описываемую квантовую систему $(|\Phi\rangle)$ можно разделить на две подсистемы, $(|\phi\rangle)$ и $(|\psi\rangle)$, и состояние этой системы $(|\Psi\rangle)$ является суперпозицией состояний двух подсистем:

$$|\Psi\rangle = \sum_{i,j} c_{ij} |\phi_i\rangle |\psi_j\rangle$$

Здесь два ортонормированных набора векторов состояния $(|\phi_i\rangle)$ и $(|\psi_j\rangle)$ описывают две части всей системы. Для такого состояния не всегда можно сказать, в каком именно состоянии находится каждая подсистема.

Note

Вспомним обсуждение кота Шредингера в [лекции про кубит](#) – упрощая до предела, можно считать, что радиоактивный атом является одной системой, а несчастный кот – второй.

До измерения ни атом, ни кот не имеют определенного состояния, а находятся в суперпозиции возможных состояний.

Что же мы можем сделать, если у нас есть доступ лишь к одной из двух подсистем, а измерить состояние второй мы уже не можем? В таком случае, все наши наблюдаемые величины будут отвечать операторами (A_ψ) , которые действует только на вторую подсистему. В примере кота Шредингера можно допустить, что экспериментально пронаблюдать мы можем лишь состояние кота, а состояние радиоактивного атома недоступно нам для измерения. В таком случае состояние второй подсистемы можно полностью описать используя редуцированную матрицу плотности. Редуцированная матрица плотности получается из матрицы плотности чистого состояния всей системы суммированием по вероятностям различных состояний первой подсистемы:

$$\rho_\psi = \operatorname{Tr}_\phi (\langle \Psi | \Psi \rangle) = \sum_{i,j} d_{ij} |\psi_j\rangle \langle \psi_i| \quad (5)$$

где

$$d_{ij} = \sum_k c_{ik} c^*_{jk}$$

Tr_ϕ означает частичный след по второй подсистеме, звездочка – комплексное сопряжение.

В таких обозначениях значение для среднего от оператора (A) вычисляется по формуле

$$\langle A_\psi \rangle = \operatorname{Tr}_\psi (A \rho_\psi)$$

Здесь след матрицы уже вычисляется по первой подсистеме.

Чаще всего при обсуждении смешанных состояний рассматривают только одну “подсистему”, считая, что вторая – это некоторый макроскопический объект (“резервуар”, например лаборатория или даже вся Вселенная). В этом случае определение матрицы плотности (1) можно рассматривать как редуцированную матрицу плотности (5), из которой убрали нерелевантные и неконтролируемые степени свободы.

Запутанные и сепарабельные состояния

Давайте вернемся к представлению состояния составной системы и зададимся вопросом: что можно сказать о связи между частями системы с точки зрения квантового описания? Для системы из двух кубитов такая составная система в общем случае может быть записана в явном виде (в этом разделе мы в основном следуем изложению из книги [01]):

$$|\Psi\rangle = a |0_A 0_B\rangle + b |0_A 1_B\rangle + c |1_A 0_B\rangle + d |1_A 1_B\rangle \quad (6)$$

где индексы (A) и (B) здесь обозначают первый и второй кубиты, соответственно, а условие нормировки дает

$$|\Psi\rangle = \sqrt{a^2 + b^2 + c^2 + d^2} = 1$$

Теперь, как можно показать, состояние типа (6) может быть представлено в виде произведения состояний двух отдельных кубитов, если $(ad = dc)$:

$$|\Psi_s\rangle = (a |0_A\rangle + b |1_A\rangle) \otimes (c |0_B\rangle + d |1_B\rangle)$$

где для выражения (6)

$$a = a_A a_B, b = a_A b_B, c = b_A a_B, d = b_A b_B$$

В других случаях, когда $(ad \neq dc)$, состояние составной системы не представимо в виде произведения

состояний подсистем, такие состояния называют несепарабельными (nonseparable). Другими словами, результат

измерения состояния подсистемы (A) будет зависеть от состояния подсистемы (B) . Это означает, что для

квантовых систем возможна нелокальная корреляция. Такое свойство квантовых систем называется

запутыванием (*entanglement*), а сами состояния запутанными (*entangled*).

Note

В отличие от английского языка, в русском языке не сложилось единой терминологии в отношении запутанных состояний. На момент написания этой лекции (осень 2021 года) в статье [Квантовая запутанность](#) русскоязычной Википедии указывается восемь (!) отличающихся терминов для этого явления, например, “запутанность”, “перепутанность” или “сцепленность”. Сделать с этим что-то сложно, остается только иметь в виду имеющиеся обстоятельства. Мы будем стараться употреблять термин **запутанность** и, соответственно, **запутанные состояния**.

Приведем пару примеров запутанных состояний:

1. $(b = c = 0, a = d = \pm 1/\sqrt{2})$ – состояние “шредингеровского кота”, см. [\[01\]](#) и [\[07\]](#), такая формула для вектора состояния возникает для суперпозиции двух макроскопически различных состояний одной из подсистем, например, живой или мертвый кот.
2. $(a = d = 0, b = -c = \pm 1/\sqrt{2})$ – такое состояние называется ЭПР-парой (EPR, от Einstein-Podolsky-Rosen) и это очень важный пример из истории изучения запутанности в квантовой физике.

Note

в 1930-е происходили многочисленные споры об “интерпретации” (сущи) квантовой механики. Именно тогда Эйнштейн, Шредингер и их коллеги обратили внимание на несопаребельные состояния и затем Эйнштейном, Подольским и Розеном был сформулирован “парадокс” – что квантовая механика либо нелокальна (т.е. несовместима с теорией относительности), либо неполна (мы учитываем не все параметры при описании состояния квантовых систем). Именно с дискуссией о сути запутанности связана знаменитая цитата Эйнштейна “Бог не играет в кости” и менее известный ответ Нильса Бора, “Альберт, не указывай Богу, что ему делать.”

Довольно долго изучение запутанности и связанных с ней трудностей считались сложным, но не основными вопросами квантовой физики. Но с развитием квантовой информатики стало понятно, что без запутанности нельзя разрабатывать квантовые компьютеры и системы квантовой связи. В настоящее время существуют устоявшиеся методы создания запутанных состояний в эксперименте. А для целей нашего курса, в симуляциях, достаточно использовать [двуухбитные гейты](#), которые обсуждались в предыдущей лекции, например **CNOT** или **CZ**, который используется в лекции про [Градиенты квантовых схем](#).

В качестве примера давайте посмотрим, как можно создать запутанное состояние в [PennyLane](#). Начнем с импортов и создания двухбитной схемы:

```
import pennylane as qml
from pennylane import numpy as np

dev = qml.device("default.qubit", wires=2)
```

Далее применим к первому кубиту операцию поворота \hat{RX} , запутаем кубиты с помощью \hat{CNOT} и далее оценим запутанность с помощью измерения оператора Паули $\hat{\sigma}_z$:

```
@qml.qnode(dev)
def circuit(param):
    qml.RX(param, wires=0)
    qml.CNOT(wires=[0, 1])
    return qml.expval(qml.PauliZ(0)), qml.expval(qml.PauliZ(1))
```

В этом примере значение переменной **param** определяет степень запутанности, и для $(\pi/2)$ запутанности будет максимальна. В результате оба кубита будут максимально смешанными и средний результат измерения будет нулевым:

```
print(circuit(np.pi / 2))
```

```
[2.22044605e-16 2.22044605e-16]
```

(код взят из официальной [демонстрации](#) для библиотеки [PennyLane](#))

Разложение Шмидта

В этом разделе мы познакомимся с важной процедурой – разложением Шмидта, которое тесно связано со спектром редуцированных матриц плотности состояния состоящей квантовой системы и благодаря которому легко увидеть, является ли состояние системы запутанным или нет.

Снова запишем чистое двухчастичное состояние $|\langle\psi\rangle\rangle$ квантовой системы, в пространстве $H_{\{\phi\}} \otimes H_{\{\psi\}}$ двух подсистем (ϕ) и (ψ) . Покажем тогда, что существуют ортонормированные состояния $|\langle i_{\{\phi\}}\rangle\rangle$ системы (ϕ) и ортонормированные состояния $|\langle i_{\{\psi\}}\rangle\rangle$ системы (ψ) , которые дадут нам разложение:

$$|\langle\psi\rangle\rangle = \sum_i \lambda_i |\langle i_{\{\phi\}}\rangle\rangle |\langle i_{\{\psi\}}\rangle\rangle, \quad (7)$$

где $\{\lambda_i\}$ – неотрицательные числа (коэффициенты Шмидта), удовлетворяющие условию $\{\sum_i \lambda_i^2 = 1\}$.

Для доказательства рассмотрим случай, когда обе подсистемы имеют пространство одинаковой размерности. Пусть тогда $|\langle n\rangle\rangle$ и $|\langle k\rangle\rangle$ состояния образуют произвольный ортонормированный базис для подсистем (ϕ) и (ψ) . Соответственно, тогда состояние системы $|\langle\psi\rangle\rangle$ может быть представлено в виде разложения:

$$|\langle\psi\rangle\rangle = \sum_n a_{nk} |\langle n\rangle\rangle |\langle k\rangle\rangle, \quad (8)$$

Константы разложения $\{a_{nk}\}$ образуют эрмитово-сопряженную комплексную матрицу (A) , которую можно привести к диагональному виду. Для этого применим к этой матрице сингулярное разложение (или SVD-разложение) вида: $(A=U \cdot S \cdot V^*)$, где (U) и (V) – унитарные матрицы, а (S) – диагональная матрица с неотрицательными действительными числами на диагонали (эти числа называют сингулярными числами матрицы (A) , и их набор однозначно определяется матрицей). Тогда разложение (8) можно привести к виду:

$$|\langle\psi\rangle\rangle = \sum_{i,n,k} s_{ii} v_{ik} |\langle n\rangle\rangle |\langle k\rangle\rangle. \quad (9)$$

Теперь переопределим базис состояний в подсистемах (ϕ) и (ψ) :

$$|\langle i_{\{\phi\}}\rangle\rangle = \sum_n u_{ni} |\langle n\rangle\rangle, \quad |\langle i_{\{\psi\}}\rangle\rangle = \sum_k v_{ik} |\langle k\rangle\rangle \quad (10)$$

и обозначим $\{s_{ii}\} \equiv \{\lambda_i\}$. В результате разложение (9) преобразуется к виду:

$$|\langle\psi\rangle\rangle = \sum_i \lambda_i |\langle i_{\{\phi\}}\rangle\rangle |\langle i_{\{\psi\}}\rangle\rangle$$

В силу унитарности (U) и (V) наборы базисных состояний $|\langle i_{\{\phi\}}\rangle\rangle$ и $|\langle i_{\{\psi\}}\rangle\rangle$ в (10) образуют полную ортонормированную систему, или базис Шмидта, а само представление (7) называют **разложением Шмидта**.

Число ненулевых значений коэффициентов Шмидта $\{\lambda_i\}$ называется числом (или рангом $\text{rank}(A)=\dim\{s_{ii} : s_{ii} > 0\}$) Шмидта для состояния $|\langle\psi\rangle\rangle$. В теории мер квантовой запутанности это число характеризует степень (или меру) запутанности состояний сложной системы. Чистое двухчастичное состояние запутанно тогда и только тогда, когда его число Шмидта > 1 , и чем больше число Шмидта, тем сильнее запутано состояние.

Следствием приведенных выше свойств является важная связь между коэффициентами Шмидта чистого запутанного состояния со спектром его редуцированных матриц плотности $\{\rho_{\{\phi\}}\} = \{\text{Tr}_{\{\psi\}}(\langle i_{\{\phi\}} | i_{\{\phi\}} \rangle)\}$ и $\{\rho_{\{\psi\}}\} = \{\text{Tr}_{\{\phi\}}(\langle i_{\{\psi\}} | i_{\{\psi\}} \rangle)\}$. Несложно убедиться, что собственные значения редуцированных матриц $\{\rho_{\{\phi\}}\}$ и $\{\rho_{\{\psi\}}\}$ совпадают и представляют собой квадраты коэффициентов Шмидта, а их собственные вектора представляют собой состояния $|\langle i_{\{\phi\}}\rangle\rangle$ и $|\langle i_{\{\psi\}}\rangle\rangle$ соответственно. Эти свойства дают нам удобный алгоритм вычисления разложения Шмидта двухчастичного состояния $|\langle\psi\rangle\rangle$ через редуцированные матрицы его подсистем: (1) на первом этапе следует вычислить редуцированные матрицы плотности $\{\rho_{\{\phi\}}\}$ и $\{\rho_{\{\psi\}}\}$; (2) на втором этапе найти общие собственные значения $\{\lambda_i\}$ и соответствующие им собственные векторы $|\langle i_{\{\phi\}}\rangle\rangle$ и $|\langle i_{\{\psi\}}\rangle\rangle$ для матриц $\{\rho_{\{\phi\}}\}$ и $\{\rho_{\{\psi\}}\}$; (3) записать разложение Шмидта в виде:

$$|\langle\psi\rangle\rangle = \sum_i \sqrt{\lambda_i} |\langle i_{\{\phi\}}\rangle\rangle |\langle i_{\{\psi\}}\rangle\rangle$$

Описания эволюции смешанного состояния

Квантовая динамика

Напомним, что квантовая динамика в терминах волновых функций $|\langle\psi\rangle\rangle$ описывается при помощи уравнения Шредингера:

$$i\hbar \frac{\partial}{\partial t} |\langle\psi\rangle\rangle = \hat{H} |\langle\psi\rangle\rangle$$

Аналогичное уравнение можно получить и для матриц плотности. Оно называется уравнением фон Неймана и записывается через коммутатор $[\hat{A}, \hat{B}]$, который определен как $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$:

$$i\hbar \frac{\partial}{\partial t} \rho = [\hat{H}, \rho]$$

Аналогично, если действие каких-то унитарных операций изменяет вектор состояния $|\langle\psi\rangle\rangle$ на $|\langle U\psi\rangle\rangle$, то матрицу плотности оно должно изменять как

$$\sum_n p_n |\langle U\psi_n\rangle\rangle \langle U\psi_n| = \hat{U}\rho\hat{U}^\dagger$$

Важное свойство унитарных матриц – их собственные значения по модулю равны единице. Действие унитарного оператора не изменяет собственных значений матрицы плотности, но вращает собственный ее базис. Исходя из этого можно сделать вывод о том, что ни чистота, ни энтропия не могут изменяться в результате унитарных операций.

Измерения и томография

Квантовая механика работает так, что любое [измерение](#) приводит к коллапсу волновой функции и является необратимым. А еще измерения, например, состояния $|\langle \text{ket}\{+\} \rangle|$ и $|\langle \text{ket}\{-\} \rangle|$ не различимы при измерениях по оси $\langle \text{mathbf}\{Z\} \rangle$ – для обоих состояний мы будем получать $|\langle \text{ket}\{1\} \rangle|$ и $|\langle \text{ket}\{0\} \rangle|$ с вероятностью 0.5 , то есть нам нужно измерять по всем базисам. В общем, получается, что восстановить амплитуду и фазу волновой функции $|\langle \text{Psi} \rangle|$ это большая проблема, если добавить сюда вероятностный характер измерения.

Note

Строго говоря это не просто “большая” проблема, а настоящая NP-полнная задача оптимизации!

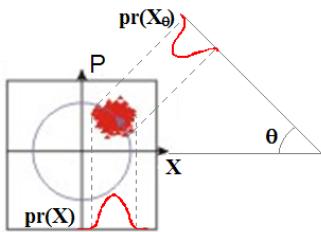


Fig. 58 Иллюстрация фазовой проблемы.

Эта задача обычно решается при помощи квантовой томографии, и восстанавливают как раз не волновую функцию $|\langle \text{Psi} \rangle|$, а матрицу плотности $|\langle \text{rho} \rangle|$ (потому что в реальных экспериментах и задачах почти не бывает чистых состояний). Представим, что наша квантовая система описывается базисом $|\text{y}_i\rangle$ – набором из (2^N) векторов, причем каждому из этих базисных векторов соответствует свое собственное значение – результат измерения (подробнее об этом было в [лекции про кубит](#)). Тогда если у нас будет достаточно много результатов измерений, то мы сможем восстановить нашу матрицу плотности $|\langle \text{rho} \rangle|$ методом максимизации правдоподобия. Выражение для правдоподобия в этом случае можно записать как:

$$L(\rho) = \prod_i |\langle \text{y}_i | \rho | \text{y}_i \rangle|^q_i$$

где q_i – это частота получения собственного значения, соответствующего волновой функции $|\langle \text{y}_i \rangle|$ (потому что измерение переводит наше состояние в базисный вектор, соответствующий результату измерения). В итоге увеличивая число измерений мы приближаем частоты q_i к вероятностям $|\langle \text{y}_i | \rho | \text{y}_i \rangle|$ ($|\langle \text{y}_i | \rho | \text{y}_i \rangle| / \sum_i |\langle \text{y}_i | \rho | \text{y}_i \rangle|$), а нашу матрицу $|\langle \text{rho} \rangle|$ к ее истинному виду.

Методы квантовой томографии являются критической частью, в том числе, квантовой связи, так как системы там обычно небольшие, но восстанавливать надо всю матрицу плотности.

Что мы узнали?

- Формализм матрицы плотности позволяет описывать составные системы (например, один кубит в многокубитной системе)
- Чаще всего в реальных экспериментах у нас ситуация “мы приготовили состояние, но точно не знаем какое” и волновая функция нам не подходит
- Что такое квантовая запутанность и как ее можно описать
- Состояние части запутанного состояния – смешанное
- Отличие смешанного состояния от чистого можно охарактеризовать параметром типа энтропии

О блоке “Фреймворки”

Этот блок включает в себя обзор фреймворков и библиотек для квантовых вычислений. Основная часть курса будет строиться вокруг библиотеки [PennyLane](#). Этот фреймворк кажется наиболее простым в освоении, а также является платформо-независимым, так как представляет собой высокоуровневый API. Дополнительные лекции этого блока расскажут также про:

- [Qiskit](#) от компании IBM;

- [cirq](#) и [Tensorflow Quantum](#) от компании Google.

Обзор фреймворков для квантовых вычислений

В этом обзоре мы кратко пробежимся по существующим сегодня фреймворкам для программирования квантовых компьютеров, а также по тем компьютерам, с которыми они совместимы.

Note

Цель этой лекции – исключительно показать как визуально выглядит тот, или иной фреймворк для программирования квантовых компьютеров. Поэтому не стоит пытаться вникнуть и понять, что именно делает тот, или иной пример приведенного тут кода! Также хочется сразу сказать, что разделение на низкоуровневые и высокоуровневые языки квантового относительно субъективно и эта лекция выражает именно точку зрения ее автора.

Квантовые ассемблеры

OpenQASM

В нашем курсе мы в основном будем пользоваться относительно высокоуровневые фреймворки. Однако это не значит, что квантовый программист не имеет возможности программировать на низком уровне. Например, компания [IBM](#) дает доступ к своему квантовому ассемблеру [OpenQASM \[CBSG17\]](#), код на котором выглядит примерно так:

```

/* Measuring the relaxation time of a qubit
 * This example demonstrates the repeated use of fixed delays.
 */
OPENQASM 3.0;
include "stdgates.inc";

duration stride = 1us;           // time resolution of points taken
const int[32] points = 50;        // number of points taken
const int[32] shots = 1000;       // how many shots per point

int[32] counts0;
int[32] counts1 = 0;  // surviving |1> populations of qubits

extern tabulate(int[32], int[32], int[32]);

bit c0;
bit c1;

defcalgrammar "openpulse";

// define a gate calibration for an X gate on any qubit
defcal x $q {
    play drive($q), gaussian(100, 30, 5);
}

for p in [0 : points-1] {
    for i in [1 : shots] {
        // start of a basic block
        reset $0;
        reset $1;
        // excite qubits
        x $0;
        x $1;
        // wait for a fixed time indicated by loop counter
        delay[p * stride] $0;
        // wait for a fixed time indicated by loop counters
        delay[p * durationof({x $1})];
        // read out qubit states
        c0 = measure $0;
        c1 = measure $1;
        // increment counts memories, if a 1 is seen
        counts0 += int[1](c0);
        counts1 += int[1](c1);
    }
    // log survival probability curve
    tabulate(counts0, shots, p);
    tabulate(counts1, shots, p);
}

```

Пример этого кода [взят](#) из официального репозитория.

PyQuil

Другим примером относительно низкоуровневого языка квантового программирования является фреймворк [PyQuil](#) [[SCZ17](#)]. Он разработан компанией [Rigetti Computing](#), производителем одноименных квантовых компьютеров и включает в себя компилятор и виртуальную машину [QVM](#). Вот так выглядит код на этом фреймворке:

```
from pyquil import get_qc, Program
from pyquil.gates import CNOT, H, MEASURE

qvm = get_qc('2q-qvm')

p = Program()
p += H(0)
p += CNOT(0, 1)
ro = p.declare('ro', 'BIT', 2)
p += MEASURE(0, ro[0])
p += MEASURE(1, ro[1])
p.wrap_in_numshots_loop(10)

qvm.run(p).readout_data['ro'].tolist()
```

Пример взят из официального репозитория.

Высокоуровневое программирование

В таких фреймворках программист имеет возможность уже не только напрямую манипулировать кубитами на низком уровне, но также вызывать высокоуровневые функции, например, [квантовое преобразование Фурье](#).

Cirq

Одним из таких фреймворков является [Cirq](#) от компании [Google](#). Код, написанный на нем можно запускать на симуляторах локально или в облаке, а также используя специально разработанные компанией [Google](#) симуляторы на базе [Tensor Processing Units](#) или [TPU](#) используя [Floq API](#). Для целей квантового машинного обучения в связке с [Cirq](#) можно использовать [Tensorflow Quantum](#) – расширение библиотеки для обычного [Deep Learning](#), которое позволяет комбинировать классические и квантовые нейронные сети вместе. Еще в экосистему [Cirq](#) входит также библиотека для квантовой химии [OpenFermion](#).

Вот так выглядит код, написанный на [Cirq](#):

```
import cirq
qubit = cirq.GridQubit(0, 0)

circuit = cirq.Circuit()
circuit.append(cirq.H(qubit))
circuit.append(cirq.measure(qubit))
print(circuit)
```

Еще больше примеров и более детальный обзор этого фреймворка читайте в нашей статье про [Tensorflow Quantum](#). Также очень рекомендуем ознакомиться с [блогом](#) “квантового” подразделения компании [Google](#) – [Google Quantum AI](#).

Q#

[Q#](#) это полноценный язык квантового программирования от компании [Microsoft](#). Этот техногигант также, как и другие, разрабатывает сейчас свои квантовые компьютеры и облачные симуляторы, а также предоставляет облачный доступ к квантовому железу других производителей. Так что программы, написанные на [Q#](#) можно запускать в облаке [Azure Quantum](#).

В отличии от большинства фреймворков, которые обычно написаны как библиотека для языка [Python](#), [Q#](#) использует [C#](#) для запуска и в целом имеет синтаксис, больше похожий на [C#](#) или [Java](#):

```

namespace QuantumRNG {
    open Microsoft.Quantum.Intrinsic;
    open Microsoft.Quantum.Measurement;
    open Microsoft.Quantum.Canon;

    @EntryPoint()
    operation GenerateRandomBits() : Result[] {
        use qubits = Qubit[4];
        ApplyToEach(H, qubits);
        return MultiM(qubits);
    }
}

```

Этот пример взят из [официальной документации](#).

На самом деле Q# это не просто язык для программирования квантовых компьютеров, а целая экосистема, в которую кроме самого языка и его стандартной библиотеки также входят:

- Библиотека [Microsoft.Quantum.Chemistry](#) для квантовой химии
- Библиотека [Microsoft.Quantum.MachineLearning](#) для квантового машинного обучения

В качестве отличного источника информации про квантовые вычисления с примерами на Q#-стеке можно использовать потрясающий репозиторий на [GitHub](#), который называется [QuantumKatas](#).

Strawberry-Fields

[Strawberry-Fields](#) это библиотека для работы с фотонными квантовыми компьютерами компании [Xanadu](#). Эти устройства являются *continuous-variable quantum computers* и в их основе лежат не бинарные состояния – кубиты, а состояния с непрерывными значениями. Учитывая сложность этого концепта, а также специфичность решаемых на таких устройствах задач мы мало будем о них говорить, но любопытный читатель может ознакомиться, например, с [этой статьей](#).

Вот так выглядит код, написанный с использованием этого фреймворка (пример из [официальной документации](#)):

```

import strawberryfields as sf
from strawberryfields import ops

# create a 3-mode quantum program
prog = sf.Program(3)

with prog.context as q:
    ops.Sgate(0.54) | q[0]
    ops.Sgate(0.54) | q[1]
    ops.Sgate(0.54) | q[2]
    ops.BSgate(0.43, 0.1) | (q[0], q[2])
    ops.BSgate(0.43, 0.1) | (q[1], q[2])
    ops.MeasureFock() | q

```

Также к [Strawberry-Fields](#) идет большое число высокоразвитых API для конкретных приложений к задачам реального мира. Например:

- [Вычисление схожести графов](#)
- [Gaussian Boson Sampling](#)
- [Поиск Clique в графах](#)
- [Поиск плотных подграфов](#)

И многие другие, с которыми можно ознакомиться на [странице](#) официальной документации.

Qiskit

[Qiskit](#) – это один из самых популярных в мире фреймворков для квантовых вычислений от компании [IBM](#). На самом деле представляет собой даже не фреймворк, а целую экосистему, в которую входят, например:

- Уже [упомянутый нами](#) квантовый ассемблер [OpenQASM](#)
- [Фреймворк](#) для программирования кубитов, написанный на [OpenQASM](#)
- [API для применения квантовых вычислений в финансах](#)
- [API для применения квантовых вычислений в ML](#)
- [API для применения квантовых вычислений в оптимизации](#)
- [API для применения квантовых вычислений в квантовой химии](#)

- [Открытый CAD для проектирования квантовых компьютеров \(!\)](#)

И [многое другое](#).

Код, написанный на [Qiskit](#) выглядит так (пример из [официальной документации](#)):

```
import numpy as np
from qiskit import QuantumCircuit, transpile
from qiskit.providers.aer import QasmSimulator
from qiskit.visualization import plot_histogram

# Use Aer's qasm_simulator
simulator = QasmSimulator()

# Create a Quantum Circuit acting on the q register
circuit = QuantumCircuit(2, 2)

# Add a H gate on qubit 0
circuit.h(0)

# Add a CX (CNOT) gate on control qubit 0 and target qubit 1
circuit.cx(0, 1)

# Map the quantum measurement to the classical bits
circuit.measure([0,1], [0,1])

# compile the circuit down to low-level QASM instructions
# supported by the backend (not needed for simple circuits)
compiled_circuit = transpile(circuit, simulator)

# Execute the circuit on the qasm simulator
job = simulator.run(compiled_circuit, shots=1000)

# Grab results from the job
result = job.result()

# Returns counts
counts = result.get_counts(compiled_circuit)
print("\nTotal count for 00 and 11 are:",counts)

# Draw the circuit
circuit.draw()
```

Код, написанный на [Qiskit](#) может быть запущен на квантовых компьютерах от компании [IBM](#). Для изучения квантовых вычислений с экосистемой [Qiskit](#) можно рекомендовать [прекрасную онлайн книгу](#). Более подробный обзор библиотеки также [в отдельной лекции нашего курса](#).

Pennylane

[Pennylane](#) это относительно высокоуровневая библиотека для квантовых вычислений и квантового машинного обучения от компании [Xanadu](#). Вот так выглядит написанный на ней код:

```
import pennylane as qml
from pennylane import numpy as np

dev = qml.device('default.qubit', shots=1000, wires=2)

def make_entanglement():
    qml.Hadamard(wires=0)
    qml.CNOT(wires=[0, 1])
    return qml.probs(wires=[0, 1])

circuit = qml.QNode(make_entanglement, dev)
circuit()
```

Эта библиотека может использовать в качестве backend любое из:

- [Cirq](#)
- [Qiskit](#)
- [Q#](#)
- [Strawberry-Fields](#)
- [PyQuil](#) (через расширение [Pennylane-Forest](#))

Но для нас самое главное преимущество этой библиотеки – это огромное количество готовых API для машинного обучения, которые позволяют автоматически считать градиенты и дифференцировать квантовые схемы, или обновлять параметры таких схем градиентным спуском.

В нашем курсе библиотеке [Pennylane](#) посвящена [целая отдельная лекций](#).

Заключение

Что мы будем использовать в курсе?

В рамках нашего курса мы в основном будем пользоваться библиотекой [Pennylane](#). Некоторые отдельные лекции будут использовать [Qiskit](#), но это скорее в порядке исключения.

Почему Pennylane?

Но почему [Pennylane](#), если самым популярным фреймворком является именно [Qiskit](#)? Приведем два основных аргумента.

1. Сегодня трудно сказать, какое квантовое железо в итоге победит, а большим преимуществом [Pennylane](#) является то, что эта библиотека является верхнеуровневой и может использовать различные backend для запуска на квантовых компьютерах разных производителей. Включая [Qiskit](#) и [IBM](#).
2. Наш курс больше про квантовое машинное обучение и именно для этих целей [Pennylane](#) подходит лучше всего. Там есть рутины и API разных уровней, позволяющих где-то самому писать градиенты квантовых схем и обновлять их параметры, а где-то просто вызывать готовый метод из [API](#), чтобы не повторять 10 раз то, что мы уже разбирали в более ранних лекциях.

PennyLane

[Pennylane](#) – библиотека Python для квантового машинного обучения, которую можно использовать для обычных квантовых вычислений. Программы, написанные на PennyLane, можно запускать, используя в качестве бэкенда настоящие квантовые компьютеры от IBM Q, Xanadu, Rigetti и другие, либо квантовые симуляторы.

Кубиты в PennyLane называются по-особому – [wires](#) (от англ. wires – провода). Такое название, скорее всего, связано с тем, что на квантовых схемах кубиты изображаются в виде продольных линий.

Последовательность квантовых операций называется [квантовой функцией](#). Такая функция может принимать в качестве аргументов только хэшируемые объекты. В качестве возвращаемого значения выступают величины, связанные с результатами измерения: ожидаемое значение, вероятности состояний или результаты сэмплирования.

Квантовая функция существует не сама по себе, она запускается на определенном устройстве – симуляторе либо настоящем квантовом компьютере. Такое устройство в PennyLane называется [device](#).

QNode

Квантовые вычисления при использовании PennyLane раскладываются на отдельные узлы, которые называются [QNode](#). Для их создания используются квантовые функции совместно с [device](#).

Создавать объекты квантовых узлов можно двумя способами: явно либо с помощью декоратора [QNode](#).

Рассмотрим первый способ – явное создание узла.

```
import pennylane as qml
from pennylane import numpy as np

dev = qml.device('default.qubit', shots=1000, wires=2)

def make_entanglement():
    qml.Hadamard(wires=0)
    qml.CNOT(wires=[0, 1])
    return qml.probs(wires=[0, 1])

circuit = qml.QNode(make_entanglement, dev)

circuit()
```

```
tensor([0.49, 0., 0., 0.51], requires_grad=False)
```

Работая с библиотекой PennyLane для математических операций, можно использовать интерфейс [NumPy](#), но при этом также пользоваться преимуществами автоматического дифференцирования, которое обеспечивает [autograd](#). Именно поэтому мы не импортировали [NumPy](#) обычным способом: `import numpy as np`, а сделали это так: `from pennylane import numpy as np`.

Второй способ создания квантовых узлов – с помощью декоратора [QNode](#). Пропускаем импорт библиотек и создание устройства, так как в начале код тот же самый:

```
@qml.qnode(dev)
def circuit():
    qml.Hadamard(wires=0)
    qml.CNOT(wires=[0, 1])
    return qml.probs(wires=[0, 1])
```

```
result = circuit()
print(result)
```

```
[0.48 0. 0. 0.52]
```

В данном примере мы взяли двухкубитную систему и создали запутанное состояние, а затем с помощью метода `probs` вычислили вероятности получения состояний $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$.

Операторы

В квантовой функции можно применять операторы X, Y, Z, S, T (`qml.PauliX`, `qml.PauliY`, `qml.PauliZ`, `qml.S`, `qml.T` соответственно), а также операторы, в которых можно задавать угол вращения вокруг одной из осей в радианах: `qml.RX`, `qml.RY`, `qml.RZ`. Здесь и далее будем использовать `qml` как псевдоним библиотеки [PennyLane](#).

В этой функции мы вращаем кубит под индексом 0 вокруг оси X на 90 градусов из начального состояния $|0\rangle$ и возвращаем [ожидаемое значение](#) `qml.PauliZ` для этого кубита с помощью `qml.expval`. Вероятности получения состояний $|0\rangle$ и $|1\rangle$ равны, так что мы получаем ожидаемое значение, близкое к нулю, что легко проверить:

$$|0.5 \cdot 1 + 0.5 \cdot (-1)| = 0$$

```
@qml.qnode(dev)
def circuit(x):
    qml.RX(x, wires=0)
    return qml.expval(qml.PauliZ(0))

circuit(np.pi/2)
```

```
tensor(0.008, requires_grad=True)
```

В следующем примере мы вращаем кубит на тот же угол 90 градусов, но уже вокруг оси Y. Ожидаемое значение в этот раз ищем для `qml.PauliX` и получаем 1, что соответствует вычислениям:

$$|1 \cdot 1 + 0 \cdot (-1)| = 1$$

```
@qml.qnode(dev)
def circuit(x):
    qml.RY(x, wires=0)
    return qml.expval(qml.PauliX(0))

circuit(np.pi/2)
```

```
tensor(1., requires_grad=True)
```

В начале этого урока мы создали устройство, которое создает и запускает одну и ту же схему 1000 раз, каждый раз производя измерения. Поменяем этот параметр:

```
dev.shots = 5
```

Посмотрим на результат каждого из этих пяти запусков и измерений для `qml.PauliZ`. Квантовая схема будет простой: применим к кубиту с индексом 1 оператор Адамара:

```
@qml.qnode(dev)
def circuit():
    qml.Hadamard(wires=1)
    return qml.sample(qml.PauliZ([1]))
```

```
circuit()
```

```
array([-1, 1, -1, 1, 1])
```

Мы получаем разные результаты: то 1, что соответствует состоянию $|0\rangle$, то -1, что соответствует состоянию $|1\rangle$.

Если вместо `qml.PauliZ` брать сэмплы для `qml.PauliX`, то результат все время будет один и тот же: 1, что соответствует состоянию $|+\rangle$ (вектор базиса Адамара).

```
@qml.qnode(dev)
def circuit():
    qml.Hadamard(wires=1)
    return qml.sample(qml.PauliX([1]))
```

```
circuit()
```

```
array([1, 1, 1, 1, 1])
```

QubitUnitary

В PennyLane можно использовать готовые операторы, либо задавать операторы явно с помощью матриц.

Для этого можно использовать класс `qml.QubitUnitary`, который принимает два параметра: `U` – квадратную унитарную матрицу и `wires` – кубиты, на которые действует оператор `U`.

В качестве примера создадим оператор, осуществляющий обмен состояний между кубитами (SWAP). Такой оператор уже есть в библиотеке PennyLane (`qml.SWAP`), но мы создадим его с помощью `qml.QubitUnitary`. Сначала мы зададим саму матрицу в виде двумерного массива, используя интерфейс NumPy:

```
U = np.array([[1, 0, 0, 0],
              [0, 0, 1, 0],
              [0, 1, 0, 0],
              [0, 0, 0, 1]])
```

Создадим заново устройство, при этом зададим число запусков схемы как `shots=1`: чтобы убедиться, что все работает правильно, нам будет достаточно одного запуска.

```
dev = qml.device('default.qubit', shots=1, wires=2)
```

Создадим и запустим схему, в которой перед применением операции SWAP, реализованной с помощью `qml.QubitUnitary`, один кубит будет находиться в состоянии 1, а другой – в состоянии 0.

```
@qml.qnode(dev)
def circuit(do_swap):
    qml.PauliX(wires=0)
    if do_swap:
        qml.QubitUnitary(U, wires=[0, 1])
    return qml.sample(qml.PauliZ([0])), qml.sample(qml.PauliZ([1]))
```

Запустим схему сначала без применения операции SWAP:

```
circuit(do_swap=False)
```

```
array([[-1],
       [ 1]])
```

А затем – с применением:

```
circuit(do_swap=True)
```

```
array([[ 1],  
      [-1]])
```

Видим, что во втором случае операция SWAP сработала: состояния кубитов поменялись местами. Можно посмотреть, как выглядит такая схема:

```
print(circuit.draw())
```

```
0: —X— U0 —| Sample[Z]  
1: ——— U0 —| Sample[Z]  
U0 =  
[[1 0 0 0]  
[0 0 1 0]  
[0 1 0 0]  
[0 0 0 1]]
```

Cirq & TFQ

Введение

[Cirq](#) – это библиотека для работы с квантовыми компьютерами и симуляторами компании *Google*. В рамках темы квантового машинного обучения нам также интересен фреймворк [Tensorflow Quantum](#) или сокращенно [TFQ](#). Это высокоуровневая библиотека, которая содержит готовые функции для квантового и гибридного машинного обучения. В качестве системы автоматического дифференцирования, а также для построения гибридных квантово-классических нейронных сетей там используется библиотека [Tensorflow](#).

⚠ Warning

Во всех дальнейших лекциях мы будем использовать в основном библиотеку [PennyLane](#), так что данная лекция исключительно обзорная и факультативная. В ней мы посмотрим несколько примеров *end2end* обучения квантовых схем на [TFQ](#) без детального объяснения теории и вообще того, что происходит. Основная цель данной лекции – исключительно обзор еще одного инструмента, а не изучение QML! Заинтересованный читатель может вернуться к этому обзору после изучения глав про [VQC. Градиенты](#) и [Квантовые нейросети](#).

Работа с кубитами

Импорты и схема

⚠ Warning

[Tensorflow Quantum](#) нельзя установить на [Windows](#), поэтому те [Makefiles](#), которые мы сделали для слушателей не содержат шаг установки [tensorflow-quantum](#). Если ваша система [POSIX](#)-совместимая и вы хотите обязательно попробовать запустить примеры кода из этой лекции, то просто в корне данного репозитория наберите команду:

```
poetry run pip install tensorflow-quantum==0.5.1
```

Обладателям компьютеров на системе [Windows](#) мы можем лишь предложить использовать [WSL2](#) или [Docker](#). В целом эта лекция факультативная и нигде далее [TFQ](#) у нас не используется.

Больше информации можно найти в [разделе про установку](#) в официальной документации этой библиотеки.

Для начала импортируем [cirq](#).

```
import cirq
```

[Cirq](#) рассчитан на работу с квантовым компьютером от компании *Google*, который представляет собой решетку кубитов. Например, вот так выглядит решетка кубитов квантового компьютера [Sycamore](#):



Fig. 59 Изображение из [AAB+19]

Поэтому в нем все строится вокруг работы с решеткой кубитов – объектом `cirq.GridQubit`. Давайте создадим кубит на решетке, который имеет координаты $((0, 0))$:

```
qubit = cirq.GridQubit(0, 0)
```

Следующей важной концепцией в `Cirq` является непосредственно квантовая схема. Давайте создадим схему, которая переводит кубит в суперпозицию состояний $(|0\rangle)$ и $(|1\rangle)$ и измеряет его:

```
circuit = cirq.Circuit()
circuit.append(cirq.H(qubit))
circuit.append(cirq.measure(qubit))
print(circuit)
```

```
(0, 0): —H—M—
```

Запуск и симуляция

Теперь создадим квантовый симулятор, который посчитает нам результат этой простой схемы на классическом компьютере:

```
sim = cirq.Simulator()
```

Как мы знаем, результат измерения такой схемы равен 50% для состояния $(|0\rangle)$, то есть если мы будем сэмплировать, то должны получать $(\text{sim} 0.5)$. Проверим это с разным числом сэмплов:

```
print(f"5 сэмплов: {sim.sample(circuit, repetitions=5).mean()}")
print(f"\n100 сэмплов: {sim.sample(circuit, repetitions=100).mean()}")
print(f"\n1000 сэмплов: {sim.sample(circuit, repetitions=1000).mean()}"
```

```
5 сэмплов: (0, 0)      0.0
dtype: float64

100 сэмплов: (0, 0)     0.48
dtype: float64

1000 сэмплов: (0, 0)    0.481
dtype: float64
```

Note

Метод `sim.sample` возвращает хорошо знакомый всем специалистам в области Data Science объект `pandas.DataFrame`.

Также у нас есть опция запустить схему через метод `run`. Может показаться, что это то же самое, но на самом деле в отличие от `sample`, метод `run` возвращает результат в несколько ином виде; а еще он позволяет запускать программу на реальном квантовом компьютере `Goolge` или их новых квантовых симуляторах на TPU:

```
print(sim.run(circuit, repetitions=25))
```

```
(0, 0)=1111110111100111001011110
```

Тут мы просто видим последовательность наших измерений.

Квантовое машинное обучение

Импорты

Мы будем использовать `Tensorflow` и `Tensorflow Quantum`.

```
import os
os.environ["TF_CPP_MIN_LOG_LEVEL"] = "3"
import tensorflow as tf
import tensorflow_quantum as tfq
```

Задача

Давайте попробуем решить игрушечную задачку классификации простой гибридной квантово-классической нейронной сетью. У нас будет один квантовый слой и один классический слой. В качестве задачи сгенерируем простенький набор данных, используя рутины `scikit-learn`. Сразу переведем входящие признаки в диапазон от нуля до π .

```
from sklearn.datasets import make_classification
import numpy as np

x, y = make_classification(n_samples=50, n_features=2, n_informative=2,
                           random_state=42, n_redundant=0)

def normalize(x):
    x_min = x.min()
    x_max = x.max()

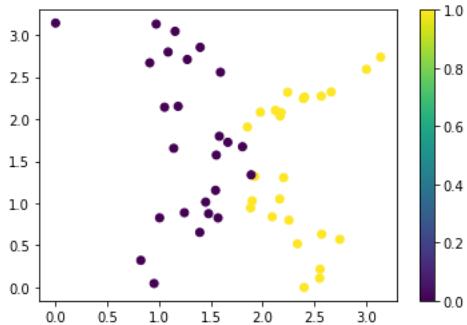
    return np.pi * (x - x_min) / (x_max - x_min)

x[:, 0] = normalize(x[:, 0])
x[:, 1] = normalize(x[:, 1])
```

Посмотрим на эти данные:

```
import matplotlib.pyplot as plt

plt.figure(figsize=(6, 4))
cb = plt.scatter(x[:, 0], x[:, 1], c=y)
plt.colorbar(cb)
plt.show()
```



Кубиты

`Tensorflow Quantum` позволяет “превращать” параметризованные схемы `Cirq` в слои нейронных сетей `Tensorflow`. Но для начала нам все равно потребуется схема. Давайте объявим пару кубитов.

```
qubits = [cirq.GridQubit(0, 0), cirq.GridQubit(0, 1)]
print(qubits)
```

```
[cirq.GridQubit(0, 0), cirq.GridQubit(0, 1)]
```

Входной слой нейронной сети

Определим входной слой, который будет кодировать наши классические данные в квантовые. Сразу закодируем данные. Так как [Tensorflow](#) работает с тензорами, то нам необходимо будет преобразовать схемы в тензор. Для этого есть специальная функция `convert_to_tensor`.

```
def data2circuit(x):
    input_circuit = cirq.Circuit()

    input_circuit.append(cirq.Ry(rads=x[0]).on(qubits[0]))
    input_circuit.append(cirq.Ry(rads=x[1]).on(qubits[1]))

    return input_circuit

x_input = tfq.convert_to_tensor([data2circuit(xi) for xi in x])
```

Слой из параметризованной схемы

Для создания параметризованных схем в [Tensorflow Quantum](#) используются символы из библиотеки символьных вычислений [sympy](#). Давайте объявим несколько параметров и создадим схему:

```
from sympy import symbols

params = symbols("w1, w2, w3, w4")

trainable_circuit = cirq.Circuit()

trainable_circuit.append(cirq.H.on(qubits[0]))
trainable_circuit.append(cirq.H.on(qubits[1]))
trainable_circuit.append(cirq.Ry(rads=params[0]).on(qubits[0]))
trainable_circuit.append(cirq.Ry(rads=params[1]).on(qubits[1]))

trainable_circuit.append(cirq.CNOT.on(qubits[0], qubits[1]))

trainable_circuit.append(cirq.H.on(qubits[0]))
trainable_circuit.append(cirq.H.on(qubits[1]))
trainable_circuit.append(cirq.Rx(rads=params[2]).on(qubits[0]))
trainable_circuit.append(cirq.Rx(rads=params[3]).on(qubits[1]))

trainable_circuit.append(cirq.CNOT.on(qubits[0], qubits[1]))

print(trainable_circuit)
```

```
(0, 0): —H—Ry(w1)—@—H—Rx(w3)—@
          |           |
(0, 1): —H—Ry(w2)—X—H—Rx(w4)—X—
```

Наблюдаемые

В качестве операторов, которые мы будем измерять, воспользуемся парой $\langle \hat{X}Y \rangle$ и $\langle \hat{Y}X \rangle$ для наших кубитов:

```
ops = [cirq.X.on(qubits[0]) * cirq.Y.on(qubits[1]), cirq.Y.on(qubits[0]) *
       cirq.X.on(qubits[1])]
```

Гибридная нейронная сеть

Теперь воспользуемся классическим [Tensorflow](#), чтобы объявить и скомпилировать нашу нейронную сеть, предварительно добавив в нее один классический слой.

- зафиксируем случайный генератор

```
tf.random.set_seed(42)
```

- входной тензор – это в нашем случае тензор типа `string`, так как это квантовые схемы

```
cirq_inputs = tf.keras.Input(shape=(), dtype=tf.dtypes.string)
```

- квантовый слой

```
quantum_layer = tfq.layers.PQC(trainable_circuit, ops)(cirq_inputs)
```

```
WARNING:tensorflow:From
/home/runner/work/qmlcourse/qmlcourse/.venv/lib/python3.8/site-
packages/tensorflow/python/ops/array_ops.py:5043: calling gather (from
tensorflow.python.ops.array_ops) with validate_indices is deprecated and will be
removed in a future version.
Instructions for updating:
The `validate_indices` argument has no effect. Indices are always validated on CPU
and never validated on GPU.
```

- классический слой и выходной слой

```
dense_layer = tf.keras.layers.Dense(2, activation="relu")(quantum_layer)
output_layer = tf.keras.layers.Dense(1, activation="sigmoid")(dense_layer)
```

- компилируем модель и смотрим, что получилось. И сразу указываем метрики, которые хотим отслеживать

```
model = tf.keras.Model(inputs=cirq_inputs, outputs=output_layer)
model.compile(
    optimizer=tf.keras.optimizers.SGD(learning_rate=0.1),
    loss=tf.keras.losses.BinaryCrossentropy(),
    metrics=[
        tf.keras.metrics.BinaryAccuracy(),
        tf.keras.metrics.BinaryCrossentropy(),
    ]
)
model.summary()
```

```
Model: "model"

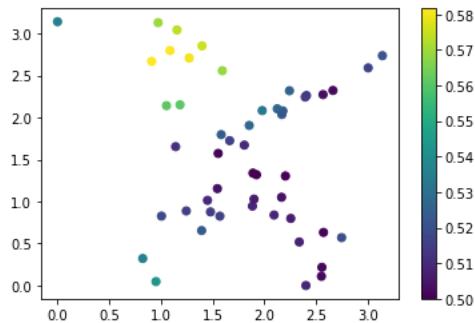
Layer (type)                  Output Shape             Param #
=====
input_1 (InputLayer)          [(None, )]              0
pqc (PQC)                     (None, 2)               4
dense (Dense)                 (None, 2)               6
dense_1 (Dense)               (None, 1)               3
=====
Total params: 13
Trainable params: 13
Non-trainable params: 0
```

Предсказания со случайной инициализацией

Наша нейросеть имеет случайные начальные параметры. Давайте посмотрим, что она предсказывает до обучения:

```
preds = model(x_input).numpy()

plt.figure(figsize=(6, 4))
cb = plt.scatter(x[:, 0], x[:, 1], c=preds)
plt.colorbar(cb)
plt.show()
```



Обучение сети

- запустим обучение

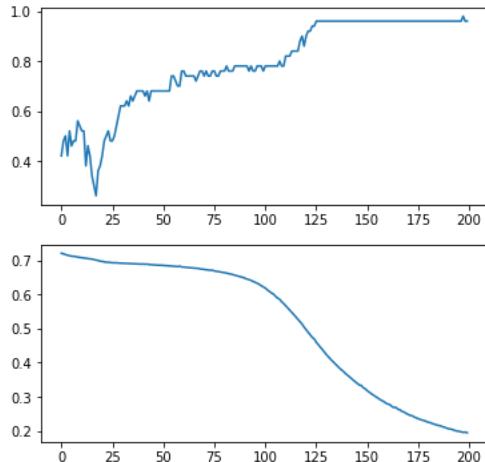
```
%time
model.fit(x=x_input, y=y, epochs=200, verbose=0)
```

```
CPU times: user 4.18 s, sys: 40 ms, total: 4.22 s
Wall time: 3.85 s
```

```
<tensorflow.python.keras.callbacks.History at 0x7efe6ca18970>
```

- визуализируем логи обучения

```
f, ax = plt.subplots(2, figsize=(6, 6))
ax[0].plot(model.history.history["binary_accuracy"])
ax[1].plot(model.history.history["binary_crossentropy"])
plt.show()
```



- визуализируем предсказания

```
preds_after_training = model(x_input).numpy()
plt.figure(figsize=(6, 4))

cb = plt.scatter(x[:, 0], x[:, 1], c=preds_after_training)
plt.colorbar(cb)
plt.show()
```



Заключение

В данной лекции мы познакомились с фреймворком [Tensorflow Quantum](#). Это достаточно мощный инструмент, особенно в связке с [Tensorflow](#), так как позволяет использовать большое число готовых методов [Tensorflow](#) и различных расширений. Тем не менее, для целей обучения [Tensorflow Quantum](#) кажется не лучшим выбором, так как имеет много неочевидного синтаксиса и предполагает, как минимум, среднего знания [Tensorflow](#). Во всех дальнейших лекциях мы будем использовать в основном библиотеку [PennyLane](#).

Qiskit

В этой лекции мы посмотрим, как при помощи библиотеки [Qiskit](#) можно работать с кубитами и квантовыми схемами.

⚠ Warning

Во всех дальнейших лекциях мы будем, за редким исключением, использовать в основном библиотеку [PennyLane](#), так что данная лекция исключительно обзорная и факультативная. В ней мы поверхностно познакомимся с концептами [Qiskit](#) и напишем пару примеров по работе с кубитами. Обзор экосистемы [Qiskit](#), включая [qiskit-machine-learning](#) выходит за рамки этой лекции, но в конце мы дадим несколько ссылок для самостоятельного изучения.

Кубиты

Импорты

```
from qiskit import QuantumCircuit
from qiskit.providers.aer import QasmSimulator

from qiskit import version
print(f"Qiskit version: {version.get_version_info()}")
```

Qiskit version: 0.18.3

Симулятор

Для запуска мы будем использовать симулятор квантового компьютера [QasmSimulator](#):

```
simulator = QasmSimulator()
```

В [Qiskit](#) поддерживается несколько симуляторов, например:

- [UnitarySimulator](#) – симулятор унитарной динамики идеальной схемы;
- [StatevectorSimulator](#) – симулятор волновой функции идеальной схемы;
- [QasmSimulator](#) – симулятор, который моделирует шумы;

Мы использовали лишь один из них. В целом, модуль [qiskit.providers](#) предоставляет базовые абстракции, которые позволяют запускать наш код на разных backend, включая реальные квантовые компьютеры, просто поменяв одну строчку в коде.

Квантовая схема

Ключевым элементом в `Qiskit` является квантовая схема – экземпляр класса `QuantumCircuit`. Создадим такую схему:

```
circuit = QuantumCircuit(2, 2)
```

В `Qiskit` необходимо явно указывать не только количество кубитов, которые нам необходимы, но также и количество классических битов, которые необходимы будут нам для измерений. В данном случае мы создали квантовую схему с двумя кубитами и двумя битами.

Квантовые гейты

Давайте попробуем воспроизвести функцию `make_entanglement` из [лекции про PennyLane](#). Применим [гейт Адамара](#) к первому кубиту, а потом применим [гейт CNOT \(CX\)](#):

```
circuit.h(0)
circuit.cx(0, 1)
```

```
<qiskit.circuit.instructionset.InstructionSet at 0x7f3e783f84f0>
```

И добавим шаг измерений:

```
circuit.measure([0,1], [0,1])
```

```
<qiskit.circuit.instructionset.InstructionSet at 0x7f3e783f86d0>
```

Компиляция и запуск схемы

Квантовые схемы `Qiskit` должны быть скомпилированы в инструкции квантового ассемблера. Для этого используется функция `transpile`:

```
from qiskit import transpile
compiled_circuit = transpile(circuit, simulator)
job = simulator.run(compiled_circuit, shots=1000)
result = job.result()
counts = result.get_counts(compiled_circuit)

print(f"Number of '00': {counts['00']}")
print(f"Number of '11': {counts['11']}")
```

```
Number of '00': 498
Number of '11': 502
```

Мы получили результат, аналогичный [полученному в лекции про PennyLane](#): вероятности получения результата $|00\rangle$ и $|11\rangle$ равны 0.5 , а другие результаты мы получить не можем.

Визуализация

В `Qiskit` есть очень мощные средства визуализации как результатов измерений, так и самих квантовых схем. Например, мы можем посмотреть результаты измерения `counts` в виде графика:

```
from qiskit.visualization import plot_histogram
plot_histogram(counts)
```



А также визуализировать саму квантовую схему:

```
circuit.draw(output="mpl")
```



В данном случае мы использовали [MatplotlibDraw](#), который позволяет делать очень красивые и легко читаемые визуализации квантовых схем – это одна из причин почему [Qiskit](#) часто используют в курсах по основам квантовых вычислений.

Экосистема Qiskit

Для более подробного изучения всей экосистемы [Qiskit](#) мы рекомендуем [онлайн книгу](#), которая также является прекрасным источником для изучения базовых и продвинутых квантовых алгоритмов. Для знакомства со средствами для квантового машинного обучения можно также изучить [документацию библиотеки qiskit-machine-learning](#), в которой представлены высокоразвитые API для таких вещей, как, например, [квантовый SVM](#), или [квантовая Generative Adversarial Network](#).

О блоке “Базовые квантовые алгоритмы”

Этот блок включает в себя:

- общий обзор наиболее известных квантовых алгоритмов;
- реализации алгоритмов Дойча и квантовой телепортации.

Продвинутые темы блока дополнительно рассказывают об алгоритмах:

- Саймона;
- Гровера.

Обзор квантовых алгоритмов

Квантовые вычисления открывают новые возможности решения задач, для которых ранее были известны только классические алгоритмы решения. С появлением идеи квантового компьютера стало понятно, что нахождение ответа для многих задач можно значительно ускорить. При этом некоторым сложным задачам, решить которые классическим способом в разумные сроки невозможно, квантовый компьютер дает реальный шанс быть решенными.

Классификация задач по временной сложности

Вообще, в соответствии с теорией алгоритмов, задачи можно разбить на классы по временной сложности их решения. Также часто используется классификация задач по объему необходимой памяти (пространственная сложность), но нас в первую очередь волнует, насколько быстро мы сможем найти правильный ответ, так что поговорим о временной сложности. Класс задач P – это те задачи, которые возможно решить на классическом компьютере за полиномиальное время, их сложность можно представить как $\mathcal{O}(n^k)$. Соответственно, решение задач этого класса не является драматически затратным по времени (за исключением случаев, когда степень полинома высока, хотя такие алгоритмы и не являются типичными).

Другой класс задач – NP (расшифровывается как “недетерминировано полиномиальные”). Это класс задач, для которых неизвестно полиномиальное классическое решение. При этом проверка решения возможна за полиномиальное время. К примеру, нахождение простых множителей большого числа с помощью известных классических алгоритмов – субэкспоненциально сложная задача (работает за время, большее, чем полиномиальное, но меньшее, чем экспоненциальное). Проверить же найденное решение предельно просто: достаточно перемножить найденные простые числа.

Среди NP -задач есть наиболее сложные задачи, которые выделяют в специальную группу под названием NP -полные задачи (или NP -complete). Если найти для них быстрое решение, то этот способ решения также можно применить к “обычным” NP -задачам.

Вообще, есть также задачи, которые, хотя и не относятся к классу NP , но несмотря на это, к ним все же можно свести задачи из NP -класса. В совокупности с NP -полными задачами они образуют класс NP -трудных задач (NP -hard).

К примеру, к NP -hard задачам относится задача коммивояжера, в которой требуется установить кратчайший путь. Как решение, так и его проверка в данном случае займут суперполиномиальное время, так что эта задача не входит в класс NP -complete. Если же ее немного упростить, так что решением будет являться путь не дольше заданного, то такая разновидность задачи является примером класса NP -complete, так проверка ее решения не требует времени, большего, чем полиномиальное.

Если дать возможность классическому компьютеру решать задачи с привнесением случайности, так что компьютер получает правильный ответ с высокой вероятностью (стандартно берут порог не менее $\frac{2}{3}$), хотя достаточно, чтобы вероятность была константой больше 0.5), то можно говорить о классе задач BPP (сокращение от англ. bounded-error, probabilistic, polynomial). Такие задачи получается решить за полиномиальное время, причем точность решения можно сколько угодно увеличивать, повторно запуская алгоритм. Квантовый аналог таких задач – класс BQP (от англ. bounded error quantum polynomial time). Это задачи, которые получается решить на квантовом компьютере за полиномиальное время, обеспечивая точность решения повторным запуском алгоритма. Наиболее известный пример такой задачи – факторизация чисел, решаемая на квантовом компьютере с помощью алгоритма Шора.

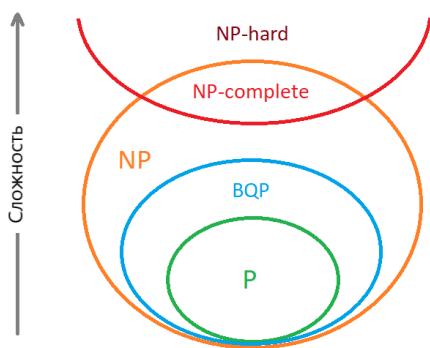


Fig. 60 Классы задач по временной сложности

На данный момент можно говорить о том, что класс BQP включает в себя в том числе и задачи, для которых неизвестно полиномиальное классическое решение, что позволяет относиться к квантовым вычислениям с оптимизмом.

Наиболее известные квантовые алгоритмы

Квантовые вычисления, несмотря на новые возможности, которые они предоставляют, все же не являются панацеей: не для всех классических “медленных” (то есть, не решаемых за полиномиальное время) алгоритмов пока удалось найти ускоренный квантовый аналог. Более того, многие даже более простые задачи в настоящий

момент выгоднее решать на классических компьютерах. Тем не менее, уже найдены квантовые алгоритмы, работающие быстрее классических. Кратко расскажем о наиболее важных из них.

Алгоритм Шора – алгоритм, наделавший больше всего шума и привлекший внимание научно-популярных СМИ к квантовым вычислениям. Действительно, этот алгоритм дает повод для беспокойства, так как он позволяет узнавать содержание сообщений, зашифрованных с помощью алгоритма шифрования RSA. Для расшифровки требуется разложить большое число на два простых множителя. Для классического компьютера решение этой задачи может занять несколько тысяч лет, а для алгоритма Шора это дело считанных часов или даже минут. Такая скорость вычислений обусловлена тем, что на квантовом компьютере удается ускорить преобразование Фурье (как прямое так и обратное). Благодаря алгоритму Шора начала развиваться квантовая криптография – шифрование, неуязвимое для атак.

Еще один алгоритм, способный преобразить мир ИТ – **алгоритм Гровера**. Благодаря ему возможно ускорить поиск по базе данных. Если на классическом компьютере решить задачу поиска элемента в базе данных возможно только перебором всех элементов, то на квантовом компьютере можно получить квадратичное уменьшение сложности, так как за счет использования эффектов суперпозиции и квантовой запутанности алгоритм Гровера “просматривает” одновременно все элементы, хотя и делает это много раз, постепенно выявляя правильное решение.

Некоторые квантовые алгоритмы пока не выглядят полезными с практической точки зрения, но даже и в таком случае они уже демонстрируют возможности, которых нет в классических вычислениях. К примеру, **алгоритм Дойча** и **алгоритм Саймона** не несут особой практической пользы в силу своей простоты, но даже такие простые примеры квантовых вычислений демонстрируют значительное ускорение (в данном случае экспоненциальное). Эти алгоритмы позволяют быстро установить свойства функций. Если алгоритм Дойча определяет, является ли функция сбалансированной, то с помощью алгоритма Саймона можно вычислить период некоторой функции.

Перспективы квантовых алгоритмов

С увеличением числа кубитов и уменьшением количества ошибок в квантовых компьютерах известные квантовые алгоритмы смогут показать себя в полной мере, но также станет возможным находить новые, более сложные и практически полезные квантовые алгоритмы. Заниматься их поиском в ближайшее время будут не только физики и математики, но и программисты, освоившие квантовые вычисления.

Алгоритмы в нашем курсе

Для более глубокого погружения в детали этих квантовых алгоритмов у нас в курсе есть отдельные лекции по:

- [Алгоритм Дойча \(начальный уровень\)](#)
- [Квантовая телепортация \(начальный уровень\) на Qiskit](#)
- [Алгоритм Гровера \(продвинутый уровень\)](#)

Алгоритм Дойча

Алгоритм Дойча (в английском варианте – **Deutsch's algorithm**) – это один из первых алгоритмов, показавших, что квантовый компьютер может решать задачи особым способом, отличающимся как от алгоритмов классического компьютера, так и от интуиции и здравого смысла человека. При этом такое решение может занимать меньшее количество шагов.

Нужно прежде всего сказать, что алгоритм Дойча не имеет практического применения в силу своей предельной простоты, зато является простейшим примером, с помощью которого можно понять, в чем состоит отличие квантовых алгоритмов от классических. Данный алгоритм был предложен в 1985 году, когда квантовых компьютеров еще не было, а практически он был реализован в 1998 году на 2-кубитном квантовом компьютере, работавшем на принципах ядерно-магнитного резонанса.

❶ Дэвид Дойч

Помимо занятий теоретической физикой в Оксфордском университете, Дэвид Дойч – автор книг “Структура реальности” и “Начало бесконечности”, в которых он популярно излагает идеи квантовых вычислений с точки зрения многомировой интерпретации (сторонником которой является) и философствует о будущем науки и человечества. Так что можно сказать, что работа алгоритма, согласно замыслу создателя, производится в параллельных вселенных. Так это или нет, пока проверить невозможно, но вычисления работают, и это главное.

Итак, в чем состоит задача, которую решает алгоритм? Представьте, что у вас есть функция, которая представляет собой “черный ящик”, принимающий на вход число из множества $\{0, 1\}$. Функция неким образом обрабатывает входное значение и возвращает число из этого же множества, то есть либо 0 , либо 1 . Нам известно, что эта функция принадлежит либо к классу сбалансированных функций, либо к классу константных функций (которые мы также можем называть несбалансированными). Задача алгоритма – установить, к какому классу принадлежит функция.

Рассмотрим все варианты этих двух классов. Всего их четыре, то есть по две функции в каждом классе. Начнем с несбалансированных:

- $f_1(x) = 0$

Это функция, всегда возвращающая 0 , независимо от входного значения. Для нее справедливы выражения:

$$f_1(0) \equiv 0$$

- $f_2(x) = 1$

Такая функция всегда возвращает 1 , то есть верно следующее:

$$f_2(0) \equiv 1$$

Ну а теперь посмотрим на сбалансированные функции. Для них характерно то, что они могут возвращать как 0 , так и 1 . В этом и заключается “баланс”.

- $f_3(x) = x$

Это тождественная функция, которая ничего не делает с входным значением. Для нее справедливо следующее:

$$f_3(0) \equiv 0$$

- $f_4(x) = \overline{x}$

А вот эта функция инвертирует входное значение, то есть возвращает не то число, которое было подано на вход, а другое:

$$f_4(0) \equiv 1$$

Классический компьютер справляется с задачей за два шага. Например, нам дана некоторая функция – “черный ящик”, и мы должны установить, сбалансирована ли она. На первом шаге мы отправляем в функцию входное значение 0 . Допустим, мы получили на выходе также 0 . Мы можем сказать, что данная функция – либо f_1 (константная функция, всегда возвращающая 0), либо f_3 (сбалансированная функция, не меняющая входное значение). Для окончательного решения мы должны сделать еще один шаг – отправить в функцию значение 1 . Если при этом мы получим опять 0 , то это функция f_1 , а если получили на выходе 1 , то искомая функция – f_3 .

Способа, с помощью которого на классическом компьютере можно за одно действие установить, сбалансирована функция или нет, не существует. И здесь свое преимущество показывает квантовый компьютер: он может установить класс функции за одно действие.

Для начала рассмотрим простейшую схему, с помощью которой можно отправлять число на вход и получать ответ от черного ящика:

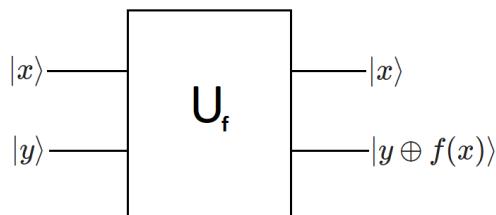


Fig. 61 Схема 1

U_f на данной схеме – это неизвестная функция (ее также часто называют “оракул”), являющаяся унитарным оператором. Обратите внимание, что в квантовой схеме используются два кубита. Это нужно для того, чтобы информация, с которой работает квантовый компьютер, не стиралась. В квантовом компьютере важно, чтобы все действия с кубитами (кроме операции измерения) были обратимыми, а для этого информация должна сохраняться. В верхнем кубите будет записано входное значение, а в нижнем – выходное значение функции. Таким образом, входное значение не будет перезаписано значением, которое вернет функция.

Но нам важно будет не только сохранить значение $|x\rangle$, но также и не разрушить $|y\rangle$. Так как кубит $|y\rangle$ очевидно имеет некоторое изначальное значение, мы не можем его просто перезаписать тем числом, которое выдаст функция $f(x)$. Здесь на помощь приходит операция исключающее ИЛИ - XOR (также ее можно называть сложением по модулю 2), обозначенная на схеме как \oplus . В процессе работы черный ящик U_f не только находит значение $f(x)$, но и применяет исключающее ИЛИ к значениям $|y\rangle$ и $|f(x)\rangle$.

Операции XOR соответствует такая таблица истинности:

a b a XOR b

0 0 0

0 1 1

1 0 1

1 1 0

Операция XOR хороша для нас тем, что она не разрушает значение $|y\rangle$, так как является обратимой.

Убедиться в этом можно, проверив тождество:

$$[(a \oplus b) \oplus b = a]$$

Схема 1 пока что не дает преимущества по сравнению с классическим компьютером, но мы можем ее немного усовершенствовать:

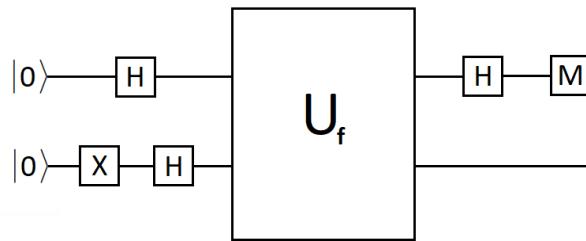


Fig. 62 Схема 2

В новой схеме оба кубита вначале будут находиться в состоянии $|0\rangle$. Затем мы применим к верхнему кубиту оператор Адамара, а к нижнему – гейт X , а затем так же, как и к верхнему, оператор Адамара. Тем самым мы приведем оба кубита в состояние суперпозиции перед тем, как они попадут на вход функции U_f . Верхний кубит будет находиться в такой суперпозиции:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

а нижний - в такой:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

Если рассмотреть это как систему $|\psi\rangle$, состоящую из двух кубитов, то она будет выглядеть так:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle)$$

Сразу после U_f система будет находиться в состоянии:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(|0\rangle + |f(x)\rangle)$$

После того, как U_f отрабатывает, нижний кубит, как это ни странно, уже нас не интересует, так что к нему операции больше не применяются, и его измерение также не производится.

Дело в том, что ответ на вопрос о том, сбалансирована функция $f(x)$ или нет, будет нами получен из верхнего кубита после того, как на него действует оператор Адамара и будет произведено измерение. В том случае, если функция сбалансирована, результат измерения верхнего кубита будет равен $|1\rangle$, а если несбалансированна – $|0\rangle$.

Разберемся подробнее, почему это происходит. Рассмотрим все возможные $f(x)$, которые могут находиться в черном ящике:

- $f(x) = f_1$

В этом случае $f(x)$ всегда принимает значение $|0\rangle$, и система кубитов будет выглядеть так:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(|0\rangle + |0\rangle) = \frac{1}{\sqrt{2}}(|00\rangle - |01\rangle + |10\rangle - |11\rangle)$$

- $f(x) = f_2$

$\langle f(x) \rangle$ будет равно $|1\rangle$ независимо от аргумента. Используя таблицу истинности XOR, легко убедиться, что во второй скобке $\langle |0\rangle$ и $\langle |1\rangle$ поменяются местами, но если вынести минус за скобку, то мы можем его не учитывать, так общий фазовый множитель (-1 в данном случае) для системы не имеет значения:

$$\langle |\psi\rangle = \frac{1}{2}(|0\rangle + |1\rangle)(|1\rangle - |0\rangle) = -\frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle) = -\frac{1}{2}(|00\rangle - |01\rangle + |10\rangle - |11\rangle)$$

Видно, что при применении функций f_1 и f_2 , являющихся несбалансированными, мы получаем фактически одно и тоже состояние. Если после этого применить к первому кубиту оператор Адамара, то после измерения мы получим значение $|0\rangle$.

Рассмотрим теперь сбалансированные функции f_3 и f_4 .

- $\langle f(x) = f_3 \rangle$

Здесь ситуация сложнее, так как $\langle f(x) \rangle$ будет зависеть от состояния первого кубита. Поэтому мы раскроем скобки, а значения функции подставим позже:

$$\langle |\psi\rangle = \frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle \oplus f(x)) - |1\rangle \oplus f(x) = \frac{1}{2}(|0\rangle |0\rangle \oplus f(x) - |1\rangle |0\rangle \oplus f(x) + |1\rangle |1\rangle \oplus f(x) - |0\rangle |1\rangle \oplus f(x))$$

$$\langle |\psi\rangle = \frac{1}{2}(|0\rangle |0\rangle \oplus 0 - |0\rangle |1\rangle \oplus 0 + |1\rangle |0\rangle \oplus 1 - |1\rangle |1\rangle \oplus 1) = \frac{1}{2}(|00\rangle - |01\rangle + |11\rangle - |10\rangle)$$

$$\langle |\psi\rangle = \frac{1}{2}(|00\rangle - |01\rangle + |11\rangle - |10\rangle) = \frac{1}{2}(|0\rangle - |1\rangle)(|0\rangle - |1\rangle)$$

Видно, что первый кубит поменял свое состояние - теперь он в суперпозиции $\langle (\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle))$, так что далее к нему можно применить оператор Адамара, после которого он перейдет в состояние $|1\rangle$.

- $\langle f(x) = f_4 \rangle$

Здесь будет похожая ситуация:

$$\langle |\psi\rangle = \frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle \oplus f(x)) - |1\rangle \oplus f(x) = \frac{1}{2}(|0\rangle |0\rangle \oplus f(x) - |1\rangle |0\rangle \oplus f(x) + |1\rangle |1\rangle \oplus f(x) - |0\rangle |1\rangle \oplus f(x))$$

$$\langle |\psi\rangle = \frac{1}{2}(|0\rangle |0\rangle \oplus 1 - |0\rangle |1\rangle \oplus 1 + |1\rangle |0\rangle \oplus 0 - |1\rangle |1\rangle \oplus 0) = \frac{1}{2}(|01\rangle - |00\rangle + |10\rangle - |11\rangle)$$

$$\langle |\psi\rangle = -\frac{1}{2}(|00\rangle - |01\rangle + |11\rangle - |10\rangle) = -\frac{1}{2}(|0\rangle - |1\rangle)(|0\rangle - |1\rangle)$$

Получили то же состояние $\langle |\psi\rangle$, что и для f_3 , с точностью до фазового множителя. Соответственно, здесь первый кубит после применения оператора Адамара также будет измерен с результатом $|1\rangle$.

Теперь можно получить более компактную формулу, которая подходит для всех четырех функций:

$$\langle |\psi\rangle = \frac{1}{2}((-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle)(|0\rangle - |1\rangle)$$

Задание

С помощью квантовых операторов попробуйте создать $\langle U_f \rangle$ для всех четырех $\langle f(x) \rangle$.

Задание рекомендуется сделать до прочтения программистской части по алгоритму Дойча, так как там содержится ответ.

Алгоритм Дойча в коде

Запограммируем алгоритм с помощью библиотеки PennyLane. Предполагается, что функция, находящаяся в черном ящике, изначально присутствует, но для учебного примера создадим также и ее, точнее, все ее четыре варианта. Для того, чтобы нам не сразу было известно, какая из этих функций анализируется алгоритмом (иначе будет неинтересно), будем использовать случайный выбор функции.

Импортируем все необходимые библиотеки и модули, а также создадим квантовое устройство-симулятор, рассчитанное на схему из двух кубитов:

```
import pennylane as qml
from pennylane import numpy as np

dev = qml.device("default.qubit", shots=1, wires=2)
```

Теперь создадим функции для черного ящика. Обратите внимание, что здесь уже учтено сложение по модулю 1 (2!) результата функции с состоянием второго кубита:

```
def f1():
    pass

def f2():
    qml.PauliX(wires=[1])

def f3():
    qml.CNOT(wires=[0, 1])

def f4():
    qml.PauliX(wires=0)
    qml.CNOT(wires=[0, 1])
    qml.PauliX(wires=0)
```

Создадим словарь с функциями и их названиями:

```
black_boxes_dict = {"f1": f1, "f2": f2, "f3": f3, "f4": f4}
```

А вот таким образом мы будем случайно выбирать название функции для черного ящика:

```
def random_black_box(black_boxes_dict):
    black_boxes_dict_list_keys = list(black_boxes_dict.keys())
    n = np.random.randint(0, len(black_boxes_dict_list_keys))

    return black_boxes_dict_list_keys[n]
```

А теперь самое важное – сам алгоритм Дойча:

```
@qml.qnode(dev, interface=None)
def circuit(black_box_name):
    qml.Hadamard(wires=0)
    qml.PauliX(wires=1)
    qml.Hadamard(wires=1)

    black_boxes_dict[black_box_name]()
    qml.Hadamard(wires=0)

    return qml.sample(qml.PauliZ([0]))
```

Итак, подготовительные действия завершены, можно приступать к демонстрации работы алгоритма.

Выберем случайным образом функцию:

```
black_box_name = random_black_box(black_boxes_dict)
```

А затем запустим алгоритм Дойча и выведем результат его работы. Собственное значение |1⟩ оператора \hat{Z} будет соответствовать состоянию $|0\rangle$ (функция несбалансированная), а собственное значение $|{-1}\rangle$ – состоянию $|1\rangle$ (функция сбалансирована):

```
result = circuit(black_box_name)
print(result)
```

1

Проверим, насколько правильно сработал алгоритм. Для этого посмотрим на функцию из черного ящика:

```
print(black_box_name)
```

f1

Также посмотрим, как выглядит квантовая схема:

```
print(circuit.draw())
```

```
0: ─H─H─| Sample[Z]  
1: ─X─H─|
```

На примере алгоритма Дойча мы видим, что уже двухкубитная схема дает прирост скорости в два раза. Если же увеличивать количество входных параметров (как в аналогичном алгоритме [Дойча-Йоки](#)), то ускорение будет экспоненциальным.

Для специалистов, занимающихся искусственным интеллектом, алгоритм может быть интересен тем, что не просто решает задачу нахождения некоторого значения, действуя как калькулятор, а дает возможность определить скрытую функцию. Это похоже на задачи машинного обучения, когда data сайентист, производя математические манипуляции с данными, в итоге получает модель (фактически – функцию), описывающую связь признаков с целевой переменной. Таким образом, интерес специалистов ИИ к квантовым вычислениям, вполне понятен, как и перспективы квантовых вычислений в этой области.

Квантовая телепортация

Пожалуй, пришло время познакомиться с квантовой телепортацией.

Часто вижу в новостях «телеported кубит», «телеported электрон». Но это лишь игра слов...

Рассмотрим два запутанных электрона – первый находится у вас, а второй у вашего друга. И под телепортацией подразумевается вот что: вы, используя допустимую операцию, изменяете состояние своего электрона, и второй (запутанный с ним), также изменит свое состояние (в соответствии с примененной операцией). Вам остается лишь выбрать допустимую операцию.

Давайте сейчас посмотрим на [состояния Белла](#) (Bell states). Прошу:

```
\ket{\psi_{00}} = \frac{1}{\sqrt{2}}(\ket{00} + \ket{11})  
\ket{\psi_{01}} = \frac{1}{\sqrt{2}}(\ket{00} - \ket{11})  
\ket{\psi_{10}} = \frac{1}{\sqrt{2}}(\ket{01} + \ket{10})  
\ket{\psi_{11}} = \frac{1}{\sqrt{2}}(\ket{01} - \ket{10})
```

Это четыре вадника квантовой запутанности. И у них есть много преимуществ: используют всего два кубита, их легко получить... А что нам еще нужно? Дальше в примере мы воспользуемся одним из этих состояний.

Теперь работаем в терминах квантовой теории информации – это сильно поможет с пониманием остальных тем нашего с вами курса.

Итак, квантовая телепортация будет проводиться двумя персонами – Алисой и Бобом, из которых первые два кубита контролируются Алисой, а третий – Бобом. Алиса хочет передать Бобу закодированное сообщение (отождествляемое с квантовым состоянием): $(|\psi\rangle = |\alpha|0\rangle + |\beta|1\rangle)$

Изначально, вся система из трех кубитов будет иметь следующее состояние: $(|\psi_0\rangle = |\psi_0\rangle|0\rangle)$. Т.е. первый кубит находится в состоянии $(|\psi_0\rangle)$, которое затем будет передаваться с помощью второго кубита на третий.

Алиса и Боб

Первые два шага являются подготовительными. Поэтому Боб находится рядом и ждет, пока Алиса проведет необходимые операции.

- Алиса применяет гейт Адамара ко второму кубиту:
$$|\psi_1\rangle = (\text{I} \otimes \text{I} \otimes H)|\psi_0\rangle|0\rangle = \frac{1}{\sqrt{2}}(|\psi_0\rangle(|0\rangle + |1\rangle))|0\rangle$$
- Далее Алиса использует гейт CNOT для того, чтобы запутать второй кубит с третьим:
$$|\psi_2\rangle = (\text{CNOT}(2,3) \otimes \text{I})|\psi_1\rangle|0\rangle = \frac{1}{\sqrt{2}}(|\psi_0\rangle(|0\rangle + |1\rangle)|0\rangle + |\psi_0\rangle|1\rangle)|1\rangle$$

Алиса и Боб

Теперь состояния готовы. Сейчас Боб, забрав третий кубит с собой, отправляется по своим делам. И Алиса, в случае необходимости, сможет передать ему послание.

- Алиса применяет CNOT между первым и вторым кубитами. Вспоминая, что $(|\psi\rangle = |\alpha|0\rangle + |\beta|1\rangle)$
$$|\psi_3\rangle = (\text{I} \otimes \text{CNOT}(1,2))|\psi_2\rangle = \frac{1}{\sqrt{2}}(|\alpha|00\rangle + |\beta|10\rangle + |\alpha|01\rangle + |\beta|11\rangle)$$
- Алиса применяет гейт Адамара на своем первом кубите:

 Внимание

Видно, что в каждой строчке у нас получаются разные состояния, по первым двум кубитам мы можем определить, в каком будет третий. Обратите внимание на общий множитель $\frac{1}{2}$ в первой строчке, не пропустите.

5. Алисе нужно теперь измерить первые 2 кубита и станет ясно, в какое из четырех состояний перейдет кубит

```

\begin{array}{l}
\hline 00 & \alpha|0\rangle + \beta|1\rangle \\
\hline 01 & \alpha|1\rangle + \beta|0\rangle \\
\hline 10 & \alpha|0\rangle - \beta|1\rangle \\
\hline 11 & \alpha|1\rangle - \beta|0\rangle
\end{array}

```

6. В зависимости от полученных по классическому каналу данных, Боб должен применить одну из операций для того, чтобы восстановить исходное состояние:

```
\begin{array}{|r|r|} \hline 00 & I \\ \hline 01 & X \\ \hline 10 & Z \\ \hline 11 & ZX \\ \hline \end{array}
```

Одним из ограничений квантовой телепортации является необходимость передать результаты измерений первых двух кубитов от Алисы к Бобу по классическому каналу (5 шаг). Поэтому весь протокол выполняется не быстрее скорости света.

Приведем данную схему:

```
from qiskit import QuantumCircuit, ClassicalRegister, QuantumRegister

# создаем необходимые регистры
qr = QuantumRegister(3, name="q")
crz = ClassicalRegister(1, name="crz")
crx = ClassicalRegister(1, name="crx")

# создаем схему
qc = QuantumCircuit(qr, crz, crx)

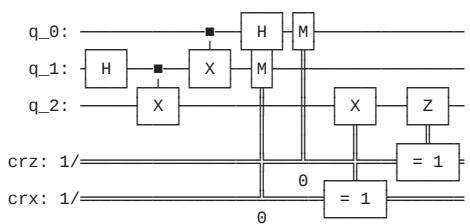
# Шаг 1
qc.h(qr[1])
# Шаг 2
qc.cx(qr[1], qr[2])

# Шаг 3
qc.cx(qr[0], qr[1])
# Шаг 4
qc.h(qr[0])

# Шаг 5 - измеряются 2 кубита Алисы, чтобы передать результат Бобу
qc.measure(qr[0], crz)
qc.measure(qr[1], crx)

# Шаг 6 - применяются гейт X и гейт Z в зависимости от того, какое из измерений дает
результат 1.
qc.x(qr[2]).c_if(crx, 1)
qc.z(qr[2]).c_if(crz, 1)

qc.draw()
```



Лекция и реализация квантовой телепортации в Qiskit

Алгоритм Гровера

Одно из самых востребованных действий в работе с данными – поиск по базе данных. При использовании классического компьютера такой поиск в худшем случае требует $\mathcal{O}(N)$ операций, где $\mathcal{O}(N)$ – количество строк в таблице. В среднем найти нужный элемент можно за $\mathcal{O}(N/2)$ операций.

фактически, это означает, что если мы не знаем, где расположен нужный элемент в таблице, то придется перебирать все элементы, пока не найдем то, что нужно. Для классических вычислений это нормально, но что, если у нас есть квантовый компьютер?

Если наша база данных работает на основе квантовых вычислений, то мы можем применить алгоритм Гровера, и тогда такой поиск потребует всего порядка \sqrt{N} действий. Конечно же, такое ускорение не будет экспоненциальным, как при использовании некоторых других квантовых алгоритмов, но оно будет квадратичным, что также довольно неплохо.



Fig. 63 Лов Гровер

Лов Гровер

Индо-американский ученый в сфере Computer Science Лов Кумар Гровер предложил квантовый алгоритм поиска по базе данных в 1996 году. Этот алгоритм считается вторым по значимости для квантовых вычислений после алгоритма Шора. Впервые он был реализован на простейшем квантовом компьютере в 1998 году, а в 2017 году алгоритм Гровера был впервые запущен для трехкубитной базы данных.

Итак, наша задача состоит в том, что мы должны найти идентификационный номер ($|Id\rangle$) записи, которая удовлетворяет определенным условиям. Функция-оракул находит такую запись (для простоты будем сначала считать, что такая запись одна) и помечает соответствующий ей $|Id\rangle$. Отметка делается достаточно оригинальным способом: $|Id\rangle$ умножается на $|-1\rangle$.

Для полной ясности соотнесем количество $|Id\rangle$ с числом кубитов в квантовой схеме. Здесь все очень просто: имея n кубитов, можно закодировать $N = 2^n$ идентификаторов. К примеру, если в таблице базы данных 1024 записей, то закодировать все $|Id\rangle$ можно с помощью десяти кубитов.

Для того, чтобы не запутаться в квантовых операциях, рассмотрим пример поменьше: с помощью двух кубитов закодируем четыре идентификационных номера, один из которых будет помечен функцией-оракулом как искомый – он будет домножен на $|-1\rangle$. Все эти четыре числа могут существовать в квантовой схеме одновременно, если кубиты приведены в состояние суперпозиции.

Пусть искомый $|Id\rangle$ равен $|11\rangle$ (будем пользоваться двоичной системой и вести счет с нуля), тогда после работы функции-оракула мы будем иметь 4 состояния: $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$. Проблема в том, что если измерить эту схему, то с равной вероятностью будет обнаружено одно из этих четырех значений, но узнать, какое из них функция-оракул пометила минусом, будет невозможно.

Получается, что одной функции-оракула недостаточно, нужно что-то дополнительное. На помощь приходит алгоритм Гровера. Правда, у него есть такая особенность – он является итерационным, то есть определенные операции (в том числе и применение функции-оракула) нужно повторить несколько раз (порядка \sqrt{N}).

Причем, с количеством итераций нельзя ошибиться, иначе алгоритм даст неправильный ответ.

В идеале после всех итераций квантовую схему можно будет измерить и получить значение $|Id\rangle$ искомой записи в таблице базы данных.

Разберем операции, которые включает в себя каждая итерация, но перед этим добавим в схему еще один кубит, который мы будем называть вспомогательным. Он нужен для хранения метки искомого индекса. Звучит не совсем понятно, но ничего сложного в этом нет, все станет ясным после разбора работы функции-оракула. Итак, наша база данных двухкубитная, но сама схема состоит из трех кубитов.

Квантовая схема выглядит так:

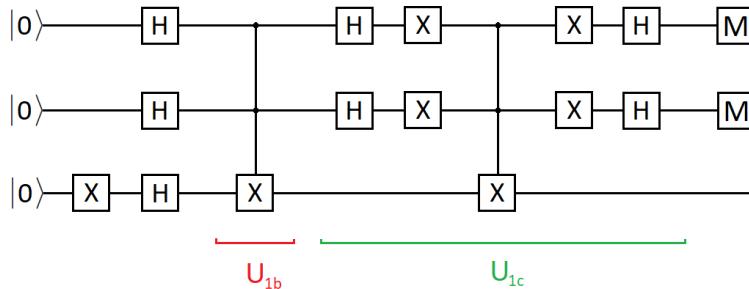


Fig. 64 Алгоритм Гровера для $n=2$ (искомый индекс 11).

В самом начале, еще до всех итераций, все кубиты (включая вспомогательный) должны быть приведены в состояние суперпозиции с помощью оператора Адамара. Причем начальное состояние всех кубитов должно быть равно $|0\rangle$, кроме вспомогательного кубита – до действия оператора Адамара он должен быть приведен в состояние $|1\rangle$.

Таким образом, вспомогательный кубит после применения оператора Адамара будет находиться в состоянии $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, тогда как остальные кубиты примут состояние $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

Далее начинаем итерации. Каждая итерация состоит из двух частей. Первая часть – это действие функции-оракула. Это некоторая функция, умеющая эффективно определять, какой индекс соответствует искомому объекту. Эта функция не может сообщить нам этот индекс напрямую, зато она может пометить его минусом.

Для разбора внутренней работы алгоритма нам потребуется задать функцию-оракул вручную и сделать ее достаточно простой, поэтому нужно знать, что в рабочих условиях она будет действовать похоже, но будет устроена, скорее всего, по-другому, так как предназначена для конкретной задачи выбора искомых данных. Мы не будем касаться вопроса конкретной реализации функции-оракула для выбора определенных данных, так как это уже другой вопрос, не влияющий на принцип алгоритма Гровера.

Для того чтобы понять алгоритм Гровера, мы должны будем понять, какие изменения происходят с состояниями кубитов до того момента, когда производится измерение, выдающее искомый индекс.

Мы договорились, что в нашей учебной задаче искомый $|11\rangle$ равен $|11\rangle$, так что в результате измерения мы должны получить именно это значение. Смоделируем оракул, который будет помечать этот индекс. В качестве такого оракула подойдет гейт Тоффоли (CCNOT). При подаче на оба его управляющих входа значений $|1\rangle$, он будет применять к управляемому кубиту (это как раз будет вспомогательный кубит) гейт $|X\rangle$.

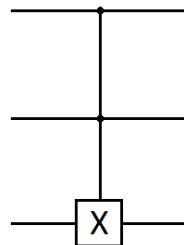


Fig. 65 Гейт Тоффоли.

На состояния верхних кубитов, кодирующих индексы $|00\rangle$, $|01\rangle$ и $|10\rangle$ гейт Тоффоли не будет реагировать, и вспомогательный кубит будет находиться в состоянии $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

Но при срабатывании гейта на индексе $|11\rangle$ к вспомогательному кубиту применится оператор $|X\rangle$, так что вспомогательный кубит примет состояние $\frac{1}{\sqrt{2}}(|1\rangle - |0\rangle)$, или, если это состояние записать по-другому, за скобками появится минус: $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

Этот минус относится не только к вспомогательному кубиту, но и ко всему состоянию, соответствующему индексу $|11\rangle$, так что можно считать, что вспомогательный кубит остался неизменным, и отнести минус к состоянию $|11\rangle$ верхних кубитов. Таким образом, индекс $|11\rangle$ помечен минусом как искомый. Другими словами, функция-оракул

перевела состояние $|11\rangle$ в состояние $|-\langle 11|q_{\text{helper}}\rangle$, где $|q_{\text{helper}}\rangle$ – вспомогательный кубит.

Запишем состояние квантовой схемы после применения оракула (состояние вспомогательного кубита – скобка справа):

$$|\psi\rangle = \frac{1}{2}\sqrt{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle)(|0\rangle - |1\rangle)$$

Итак, первая часть первой итерации завершена. Искомый индекс помечен, но если измерить кубиты прямо сейчас, то это ничего не даст – минус не проявится при измерении. Да и сам индекс $|11\rangle$ будет получен только с вероятностью (0.25) – такой же, как и других индексов.

Для того, чтобы лучше понять дальнейшие действия, представим первую половину работы алгоритма в виде рисунка, показывающего вектор текущего состояния. В качестве единичного вектора горизонтальной оси мы будем использовать все состояния из суперпозиции кроме того, который соответствует искомому индексу, а вертикальным единичным вектором будет искомый вектор.

Вектор $|c\rangle$ – состояние системы перед первой итерацией – является линейной комбинацией векторов, соответствующим горизонтальной и вертикальной осям.



Fig. 66 Состояние системы перед первой итерацией.

Можно выразить вектор $|c\rangle$ для нашего случая (системы из двух кубитов с искомым индексом $|11\rangle$), обозначив его координаты за $|x\rangle$ и $|y\rangle$:

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = x \frac{1}{\sqrt{3}}(|00\rangle + |01\rangle + |10\rangle) + y |11\rangle$$

Из данного уравнения находим $x = \frac{\sqrt{3}}{2}$ и $y = \frac{1}{2}$.

По этим координатам можно понять, что угол между вектором $|c\rangle$ и горизонтальной осью (обозначим его как $\langle\theta\rangle$) равен $\langle\frac{\pi}{6}\rangle$. Если забежать немного вперед, то можно сказать, что наша цель – добиться, чтобы текущее состояние дошло до $\langle\frac{\pi}{2}\rangle$ (или хотя бы приблизительно), то есть почти или полностью равнялось искомому состоянию, так что после измерения можно было его и получить с высокой вероятностью.

Координаты текущего вектора состояния можно записать через угол $\langle\theta\rangle$:

$$\begin{cases} x \equiv \cos(\theta) \\ y \equiv \sin(\theta) \end{cases}$$

На всякий случай нужно уточнить, что вспомогательный кубит не отражается на рисунке с окружностью, так как он не предназначен для обозначения индекса, а только хранит в себе его метку.

После применения функции-оракула текущий вектор отразится относительно горизонтальной оси. Объясняется это очень легко – его вертикальная компонента (вектор $|11\rangle$) становится отрицательной.

Вектор $|c_{1b}\rangle$ – это отражение вектора $|c\rangle$ на угол $\langle\theta\rangle$ вниз относительно горизонтальной оси:

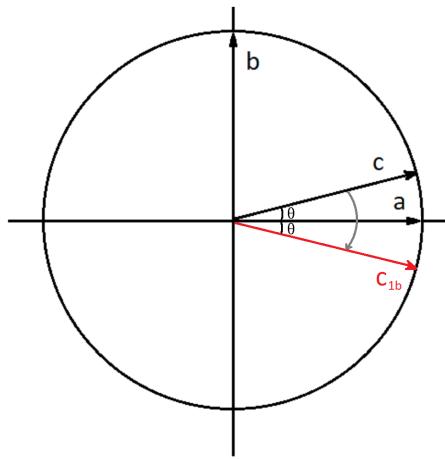


Fig. 67 Состояние системы после первой части первой итерации.

Такое отражение в нашем примере производится с помощью операции CCNOT , но в общем случае операция выглядит так:

$$\langle U_{1b} = I - 2|b\rangle\langle b|$$

Функцию-оракул мы здесь обозначили как $\langle U_{1b} \rangle$. Она меняет знак только для вертикальной составляющей вектора состояния, поэтому и происходит отражение.

Проверим формулу в действии, применив ее для нашего примера:

$$\langle U_{1b} |c\rangle = (I - 2|11\rangle\langle 11|) \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle)$$

И наконец приступаем к разбору второй части первой итерации. В ней будет происходить еще одно отражение вектора, но уже не относительно горизонтальной оси, а относительно вектора $\langle c \rangle$. Нетрудно заметить, что при этом текущий вектор состояния станет равен $\langle (\cos 3\theta)|a\rangle + \langle \sin 3\theta|b\rangle \rangle$.

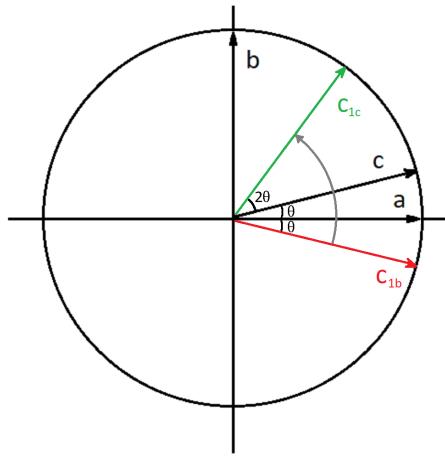


Fig. 68 Состояние системы после второй части первой итерации.

Операция для получения вектора $\langle c_{1c} \rangle$ будет выглядеть так:

$$\langle U_{1c} = 2|c\rangle\langle c| - I \rangle$$

Посчитаем, чему равен вектор $\langle c_{1c} \rangle$ для нашего примера:

$$\langle U_{1c} |c_{1b}\rangle = (2|c\rangle\langle c| - I) \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle) = |11\rangle$$

Произошло отражение текущего вектора состояния $\langle |c_{1b}\rangle$ относительно вектора $\langle |c\rangle$. Если представить $\langle |c_{1b}\rangle$ как $\langle k_1 |c\rangle + \langle k_2 |b\rangle$, где $\langle |c_{1b}\rangle$ – вектор, перпендикулярный $\langle |c\rangle$, а $\langle k_1 \rangle$ и $\langle k_2 \rangle$ – действительные коэффициенты, то тогда отраженный вектор будет равен $\langle k_1 |c\rangle - \langle k_2 |b\rangle$.

В нашей квантовой схеме эта часть итерации реализована таким образом:

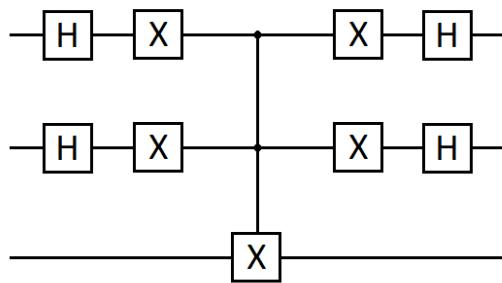


Fig. 69 Вторая часть итерации.

Вначале применяется оператор Адамара для первого и второго кубитов. Это упрощает нашу задачу, так как теперь отразить вектор состояния нужно не относительно состояния суперпозиции $(\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle))$, а относительно состояния $(|00\rangle)$.

Далее требуется отразить вектор, то есть всем состояниям кроме нулевого присвоить минус, но мы сделаем проще: присвоим минус нулевому состоянию $(|00\rangle)$, а остальные состояния, составляющие суперпозицию, оставим как есть. Сделаем это, последовательно применив гейты (X) и $(CCNOT)$. После этого вернемся в исходную "систему координат", применив операции в обратном порядке: сначала (X) , а потом (H) .

Из-за применения такого лайфхака (присвоения минуса нулевому состоянию) мы в нашем двухкубитном примере получим результат с точностью до общей фазы: не $(|11\rangle)$, а $(-|11\rangle)$. Но это не страшно, так как после измерения мы все равно увидим искомое значение индекса.

По рисунку, изображающему на окружности состояние системы после второй части первой итерации, видно, что в общем случае каждая последующая итерация будет приближать текущий вектор к вертикальному. Но в нашем случае угол между вектором состояния и горизонтальной осью после окончания первой итерации равен (3θ) , то есть это уже и есть желаемый угол $(\frac{\pi}{2})$.

В общем случае этот угол равен $((2t + 1)\theta \approx \frac{\pi}{2})$, где (t) – номер произведенной итерации. Отсюда можно вывести число итераций, необходимое для работы алгоритма. При большом требуемом количестве итераций (t) и фиксированном $(K = 1)$ (для $(K > 1)$ вывод формулы аналогичный) угол (θ) будет близок к (0) , так что его можно заменить на $(\sin(\theta))$, который, в свою очередь, равен $(\frac{1}{\sqrt{N}})$:

$$[(2t + 1)\theta \rightarrow (2t + 1)\frac{1}{\sqrt{N}} \approx \frac{\pi}{2}]$$

Если пренебречь единицей в скобках на основании того, что (t) – большое число, можно найти, что (t) приблизительно равно $(\frac{1}{\sin(\theta)} \sqrt{N})$.

Мы уже разобрались, что каждая итерация состоит из двух частей. Первая часть – отражение вниз относительно горизонтальной оси. Вторая часть – отражение вверх относительно изначального состояния, то есть вектора (c) . Вектор состояния всегда будет отражаться вверх на больший угол, чем в первой части итерации. Этим и будет обеспечиваться его постепенное приближение в вертикальной оси.

Мы разобрали случай, когда требуется найти один объект в таблице. Если же потребуется найти несколько объектов, то тогда, обозначив их количество за (K) , мы должны будем проделать около $(\frac{K}{4}\sqrt{N})$ итераций. Таким образом, для успешной работы алгоритма Гровера необходимо знать число (K) , чтобы можно было найти через него угол (θ) , а затем число итераций.

Реализация алгоритма Гровера

Итак, мы разобрали общие принципы алгоритма Гровера, а также рассмотрели конкретный пример. Настало время написать для этого примера программу.

Для начала импортируем все необходимые библиотеки и создадим схему из трех кубитов:

```
import pennylane as qml
from pennylane import numpy as np

dev = qml.device('default.qubit', shots=1, wires=3)
```

Начальная функция, создающая суперпозицию для каждого кубита:

```

def U_start():
    qml.PauliX(wires=2)
    for i in range(3):
        qml.Hadamard(wires=i)

```

Создадим функцию, действующую аналогично оракулу (первая часть итерации). Эта функция помечает значение индекса $|11\rangle$:

```

def U_b():
    qml.Toffoli(wires=[0, 1, 2])

```

Вторая часть итерации:

```

def U_c():
    for i in range(2):
        qml.Hadamard(wires=i)
        qml.PauliX(wires=i)

    qml.Toffoli(wires=[0, 1, 2])
    for i in range(2):
        qml.PauliX(wires=i)
        qml.Hadamard(wires=i)

```

Объединим первую и вторую часть итерации в одну функцию:

```

def U_iteration():
    U_b()
    U_c()

```

Переходим к итоговой функции, содержащей все шаги, а также производящей измерение кубитов в конце. В аргументе $|N\rangle$ мы должны будем указать количество итераций:

```

@qml.qnode(dev)
def circuit(N: int):
    U_start()
    for t in range(N):
        U_iteration()
    return qml.probs(wires=[0, 1])

```

Запускаем функцию и выведем ее результат:

```

print(circuit(N=1))

```

```
[0. 0. 0. 1.]
```

Так как в качестве искомого индекса выступало значение $|11\rangle$, то в результате запуска функции мы должны получить массив, состоящий из вероятностей каждого индекса, в котором искомый индекс (в нашем примере он будет последним в массиве) должен иметь наибольшую вероятность. Параметр устройства `shots` при необходимости можно увеличивать, не забывая о том, что его увеличение будет кратно замедлять алгоритм. Таким образом, мы нашли с помощью алгоритма Гровера искомый индекс.

Алгоритм Гровера может применяться не только для задач простого поиска в базе данных, но и как дополнительное средство ускорения для поиска экстремума целочисленной функции, а также для поиска совпадающих строк в базе данных, так что этот алгоритм, как и его модификации, сможет быть полезным в разнообразных задачах Data Science.

Задание

1. Распишите операторы (U_{1b}) и (U_{1c}) из примера в виде матриц (4×4) и проведите расчеты для получения (c_{1b}) и (c_{1c}) в виде векторов-столбцов.
2. Модифицируйте приведенный выше код алгоритма Гровера для двухкубитной базы данных так, чтобы искомый индекс соответствовал состоянию $(|00\rangle)$.

Алгоритм Саймона

Задача Саймона

Давайте начнём с того, что алгоритм Саймона решает задачу Саймона. Да, вот такой замечательный учёный - нашёл проблему, решил проблему.

По своей природе задача Саймона является разновидностью задачи о скрытой абелевой подгруппе [Lom04].

Пусть задана функция $f: \{0, 1\}^n \rightarrow \{0, 1\}^n$ и **неизвестная** строка $s \in \{0, 1\}^n$, для всех $(x, y) \in \{0, 1\}^n$ выполняется:

\[\large f(x) = f(y) \Leftarrow

То есть, если мы для двух различных строк $\setminus(x)$ и $\setminus(y)$ имеем одинаковое значение $\setminus(f(x) = f(y))$, то $\setminus(x \oplus y)$

равняется некоторой неизвестной строке `\(s)`. Функция `\(f(x)\)` представляет собой чёрный ящик.

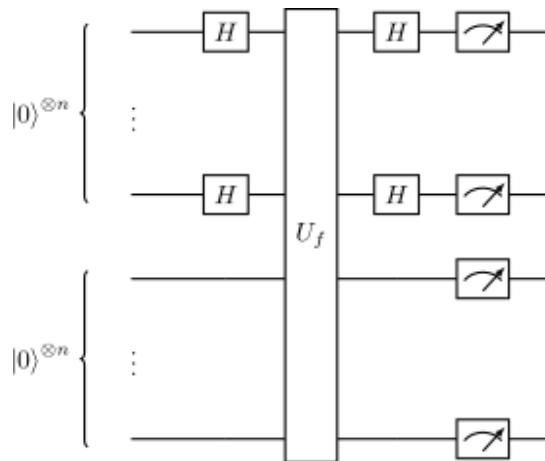
Задача состоит в том, чтобы найти $\ell(s)$, выполнив при этом как можно меньшее количество вызовов $\ell(f(x))$.

Описание алгоритма

Сразу определим операцию на двух бинарных строках (они же бинарные векторы) $\langle x = x_0 \ x_1 \ x_2 \dots x_{\{2^n-1\}} \rangle$, $\langle z = z_1 \ z_2 \ z_3 \dots z_{\{2^n-1\}} \rangle$:

$$\| \langle x, z \rangle \| = \bigoplus_{i=0}^{2^n-1} x_i \wedge z_{-i} = x_0 \wedge z_0 \oplus x_1 \wedge z_1 \oplus \dots \oplus x_{2^n-1} \wedge z_{2^n-1} \|$$

Принципиальная схема алгоритма Саймона:



В принципе, есть небольшие вариации в реализации алгоритма, мы рассмотрим наиболее простую (все необходимые ссылки в конце приведены):

1. Сначала приготовления. Вначале мы приготовим 2 набора квантовых регистров (каждый размерности $\lceil n \rceil$) в следующем состоянии:

$$|\psi_1\rangle = |\text{orange}\rangle |\text{orange}\rangle$$

2. Применяем гейт Адамара на первом регистре:

$$|\text{H}^{\otimes \lceil n \rceil} \psi_1\rangle = |\psi_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

3. Применяем оператор (U_f) , который выполняет следующее преобразование $(U_f)|x\rangle = |\psi_2\rangle = \frac{1}{\sqrt{2}}(|f(x)\rangle + |1-f(x)\rangle)$. Во втором регистре записывается значение функции $f(x)$:

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \sum_x |x\rangle |f(x)\rangle$$

4. Производим измерение на втором регистре. Измеренное значение будет соответствовать либо $|0\rangle$ либо $|1\rangle$

$$\langle \psi_3 | = \langle \text{orange} | \langle \text{orange} | \psi_2 \rangle = \langle \text{orange} | \langle f(x) | \psi_2 \rangle = \langle \text{orange} | \langle f(x) | \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \sum_x |x\rangle |f(x)\rangle = \frac{1}{\sqrt{2}}(\langle \text{orange} | |0\rangle + \langle \text{orange} | |1\rangle) \sum_x \langle x | |f(x)\rangle$$

5. Снова применяем гейт Адамара на первом регистре:

$$|\text{H}^{\otimes \lceil n \rceil} \psi_3\rangle = |\psi_4\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \sum_z (-1)^{\langle z | x \rangle} |z\rangle$$

(операцию $\langle z | x \rangle$ мы уже определяли выше).

 Note

Для чисел $\forall x \exists y \neg 110111 \wedge 11 \vee \forall z \neg 010101$ получим $\neg \wedge 0 \oplus \wedge 1 \oplus \wedge 0 \oplus \wedge 1 \oplus \wedge 1 \oplus \wedge 0 \oplus \wedge 1 = 1$

6. Производим измерение на первом регистре. И здесь возможны 2 варианта исхода:

- $$1. \quad x + y = s = 0^n$$

Учитывая $\langle x_1 \oplus x_2 = 0 \rangle$ и то, что состояние всей системы описывается как $|f(x) \rangle |x\rangle$

$|x\rangle$

находим вероятность получить строку $\langle z \rangle$ на первом регистре

$|\frac{1}{2}|^2$ или $|\frac{1}{2}|^2$

т.е. в данном случае имеет место **равномерное распределение**.

2. $\langle x_1 \oplus x_2 = s \rangle$

Этот случай гораздо интереснее. Функция $|f\rangle$ преобразует два различных входных значения $\langle x_1, x_2 \rangle$ в одно $\langle f(x_1) = f(x_2) = s \rangle$ в $\langle 0, 1 \rangle^n$. Также $\langle x_1 \oplus x_2 = s \rangle$ эквивалентно $\langle x_1 \oplus x_2 = 3 \rangle$

$$\begin{aligned} |\frac{1}{2}|^2 &= \frac{1}{2^n} \sum_{z \in \{0, 1\}^n} \sum_{x \in \{0, 1\}^n} \frac{(-1)^{\langle z \mid x \rangle}}{\langle f(x) \rangle} \langle x \rangle^2 = \\ &= \frac{1}{2^n} \left(\sum_{x \in \{0, 1\}^n} \sum_{z \in \{0, 1\}^n} \frac{(-1)^{\langle z \mid x \rangle}}{\langle f(x) \rangle} \langle x \rangle^2 \right) = \\ &= \frac{1}{2^n} \left(\sum_{x \in \{0, 1\}^n} \sum_{z \in \{0, 1\}^n} \frac{(-1)^{\langle z \mid x \rangle}}{\langle f(x) \rangle} \langle x \rangle^2 \right) = \\ &= \frac{1}{2^n} \left(\sum_{x \in \{0, 1\}^n} \sum_{z \in \{0, 1\}^n} \langle x \rangle \langle (-1)^{\langle z \mid x \rangle} |f(x)\rangle \langle x \rangle^2 = \right. \\ &\quad \left. \langle z \rangle \langle (-1)^{\langle z \mid x \rangle} |f(x)\rangle \langle x \rangle^2 \right) = \\ &= \frac{1}{2^n} \left(\sum_{z \in \{0, 1\}^n} \langle z \rangle \langle (-1)^{\langle z \mid x_1} |f(x_1)\rangle + (-1)^{\langle z \mid x_2} |f(x_2)\rangle \langle z \rangle \right) = \\ &= \frac{1}{2^n} \left(\sum_{z \in \{0, 1\}^n} \langle z \rangle \langle (-1)^{\langle z \mid x_1} |f(x_1)\rangle + (-1)^{\langle z \mid x_2} |f(x_2)\rangle \langle z \rangle \right) = \end{aligned}$$

Расчёты вероятностей можно найти в приложениях [Приложение 1](#).

Выполняем алгоритм n раз. После чего у нас будет система n линейно независимых уравнений. Теперь приступаем к вычислению строки s .

7. Постобработка.

Итак, для того, чтобы найти $s = (s_0, s_1, s_2, \dots, s_{n-1})$, нам потребуется n линейно независимых векторов $\langle \vec{z}_i \rangle$, для которых выполняется $\langle \vec{z}_i, s \rangle = 0$.

После того как получена система из n линейно независимых уравнений, решение можно найти методом Гаусса.

Пример

Давайте возьмём $n = 3$, строку $s = 100$, и функцию f , которая соответствует критерию

$f(x) = f(y) \Leftrightarrow x \oplus s = y$

Обычно функция $|f(x)\rangle$ задана наперёд. Ну а мы выберем её простейшей: $|f(x)\rangle = x \oplus s$.

Давайте посмотрим на таблицу истинности всех нужных переменных.

$$\begin{array}{c|ccc|ccc|ccc|ccc} & \text{000} & \text{001} & \text{010} & \text{011} & \text{100} & \text{101} & \text{110} & \text{111} \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{array}$$

Нарисуем схему на Qiskit, которая будет показывать одну итерацию алгоритма:

```
from qiskit import QuantumCircuit, ClassicalRegister, QuantumRegister

# Работаем в пространстве размерности n = 3.
n = 3

# Создаём необходимые регистры
qr1 = QuantumRegister(n, name="q1")
qr2 = QuantumRegister(n, name="q2")
cr1 = ClassicalRegister(n, name="c1")

# Шаг 1. Создаём квантовую схему (инициализация состояния)
qc = QuantumCircuit(qr1, qr2, cr1)

# Шаг 2. Применяем гейт Адамара ко всем кубитам первого регистра
qc.h(range(n))

# Шаг 3. Применяем U_f
qc.cx(qr1[0], qr2[0])

# Шаг 4. Производим измерение первого регистра
qc.measure(qr2, cr1)

# Шаг 5. Ещё раз применяем гейт адамара к каждому из кубитов
qc.h(range(n))

# Шаг 6. Производим измерение первого регистра
qc.measure(qr1, cr1)

# Рисуем схему
qc.draw()
```



Теперь пройдём по всем шагам алгоритма:

- Инициализация всех регистров в |000> состоянии:

$$|\psi_0\rangle = |\psi_0\rangle_{\{1\}} \otimes |\psi_0\rangle_{\{2\}}$$
 - Применяем Адамар к первому регистру:

$$|\psi_1\rangle = |\psi_1\rangle_{\{1\}} = \frac{1}{\sqrt{2}}(|000\rangle + |001\rangle + |010\rangle + |011\rangle) + |100\rangle + |101\rangle + |110\rangle + |111\rangle$$
 - Применим оракул U_f:

$$U_f(|\psi_1\rangle) = |\psi_2\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |001\rangle + |010\rangle + |011\rangle + |100\rangle + |101\rangle + |110\rangle + |111\rangle - |100\rangle - |101\rangle - |110\rangle - |111\rangle)$$
 - Измеряем второй регистр. С вероятностью $\frac{1}{2}$ мы получим либо значение |000> либо |100>. Допустим результат равен |100>. Тогда состояние первого регистра примет вид: $|\psi_3\rangle = (|100\rangle + |101\rangle + |110\rangle + |111\rangle)$.
 - Ещё один раз применяем бейты Адамара на первый регистр:

$$|\psi_4\rangle = \frac{1}{\sqrt{4}}(|000\rangle + |001\rangle + |010\rangle + |011\rangle + |100\rangle + |101\rangle + |110\rangle + |111\rangle)$$
 - Производим измерение первого регистра, записывая результат.
 Выполняем шаги 1-6 n раз.
 - Здесь решаем систему уравнений.

Приложение

```
(1)^{\langle z \rangle} = \begin{cases} 1 & \text{если } z = 0 \\ -1 & \text{если } z = 1 \end{cases}
```

Ссылки

- [Simon Algorithm](#)
- [Simon's problem](#)
- [Qiskit Simon algorithm](#)

О блоке “Железо квантовых компьютеров”

Этот блок включает в себя обзор различных подходов к физической реализации квантовых вычислений. Как и классические компьютеры, квантовые компьютеры вовсе не обязаны быть основаны на какой-то одном физическом принципе работы: информация может храниться в токах и напряжениях в электрической схеме, спинах атомов, механическом положении или движении микроскопических частиц. Квантовая механика проявляется для любых достаточно малых систем. В качестве критерия малости можно использовать постоянную Планка \hbar , которая имеет размерность фазового объема, или произведения координаты на импульс; это отражается и в принципе неопределенности Гейзенберга. Это значит, что если мы имеем систему, работающей на частоте ω , разность энергий уровней, в которых хранится информация, скорее всего будет порядка $\hbar\omega$. Из-за того, что $\hbar\omega \sim 1.054 \times 10^{-34} J \cdot s$ – очень маленькое число, два квантовых состояния очень слабо отличаются друг от друга.

Из-за того, что все платформы для квантовых вычислений очень разные, делать конкретные утверждения про все сразу практически невозможно. В этом блоке будет дана обзорная классификация разных платформ по разным признакам со ссылками на оригинальные работы, а затем более подробная лекция про сверхпроводниковые кубиты.

Обзор и классификация физических платформ для квантовых вычислений

Радиочастотные и оптические кубиты

Первый признак, по которому можно классифицировать разные кубиты, это характер тех двух квантовых состояний, в которых хранится информация. Рассмотрим, например, фрагмент структуры уровней для иона Yb-171+ [Mon]. Этот ион часто используют для квантовых вычислений.

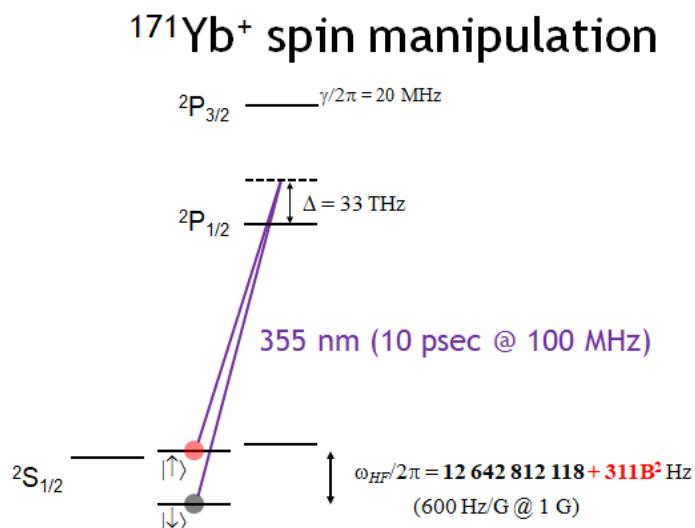


Fig. 70 Фрагмент структуры энергетических уровней иона Yb-171+. У типичного атома очень много степеней свободы: спин ядра, спин электронной подсистемы, радиус орбитали электрона и проекция момента орбитали электрона. Переходы между состояниями имеют очень разные частоты, времена когерентности и восприимчивости к частотам к внешним полям. Чтобы сделать из иона кубит, надо определиться, какой уровень будет “0”, а какой – “1”.

На сегодняшнем этапе все платформы для квантовых вычислений можно разделить на два класса: те, которые работают со сравнительно низкими $\omega/2\pi \sim 10^{10} - 10^{15}$ Гц частотами (вплоть до микроволновых), и те, которые работают на высоких частотах $\omega/2\pi \sim 10^{13} - 10^{15}$ Гц. К первым относятся:

- сверхпроводниковые квантовые схемы;
- кубиты на тонком и сверхтонком расщеплении в нейтральных атомах;
- кубиты на сверхтонком расщеплении в ионах;
- кубиты спинах в квантовых точках;
- кубиты на экзотических квазичастицах, вроде Майорановских фермионов.

К второй группе относятся:

- оптические и телекоммуникационные фотона;
- кубиты на оптических переходах в атомах и ионах.

Квантовые компьютеры первого типа надо сильно охлаждать, так как они самовозбуждаются под действием броуновского движения и теплового излучения тел, находящихся при комнатной температуре.

Природа степени свободы для кодирования информации

Можно классифицировать популярные платформы для квантовых вычислений и по другим признакам. Например, есть кубиты на естественных степенях свободы и кубиты на коллективных степенях свободы.

Естественные степени свободы – это:

- любые переходы в уединенных атомах и ионах;
- фотоны в открытом пространстве.

Кубиты на естественных степенях свободы автоматически получаются одинаковыми: вы берете одинаковые атомы, и у них обязательно будут одинаковые переходы. Впрочем, если поместить их в какую-нибудь ловушку, то из-за того, что ловушки немного влияют на частоты переходов в атомах, эту эквивалентность легко испортить.



Fig. 71 Фотография уединенного атома в ловушке. Источник: [\[ion\]](#)

Кубиты на искусственных степенях свободы получаются когда имеет место коллективная динамика большого числа частиц, например поляризация в материала оптоволокна, или коллективное движение электронов в сверхпроводнике. Примеры искусственных степеней свободы:

- сверхпроводниковые электрические цепи;
- фотоны в интегрально-оптической схеме;
- фотоны в оптоволокне;
- спины в квантовых точках;

- любые квазичастицы в твердотельных системах.



Fig. 72 Изображение кремниевой микросхемы с кубитом на квантовой точке. Источник: [SZK+18]

На самом деле граница между естественными степенями свободы и искусственными вовсе не такая очевидная, как может показаться на первый взгляд. Так, например, фотон, находящийся в открытом пространстве, можно поместить между двумя зеркалами, которые образуют резонатор. Но в таком случае в зеркалах, и в носителях заряда в зеркале будет храниться часть энергии фотона – получается так называемый “поляритон”, который уже не совсем естественная степень свободы. Абсолютно тождественными могут быть только те носители информации, которые ни с чем не связаны.

Способ кодирования информации

Только системы, где есть всего два уровня, а больше никакие не используются и не учитываются, строго говоря являются кубитами. Такими являются, например:

- потоковые сверхпроводниковые кубиты;
- логические кубиты, которые получаются выполнением кодов коррекции на большом числе физических кубитов любой природы.

Помимо “чистых” кубитов, у которых есть всего два дискретных состояния, существуют также “кубиты”, которые имеют \sqrt{d} ($d > 2$) состояний. Таковыми являются, например:

- ионы со сложной структурой уровней и правилами отбора переходов между ними;
- сверхпроводниковые кубиты-трансмопы;
- атомы и ридберговские состояния в них.

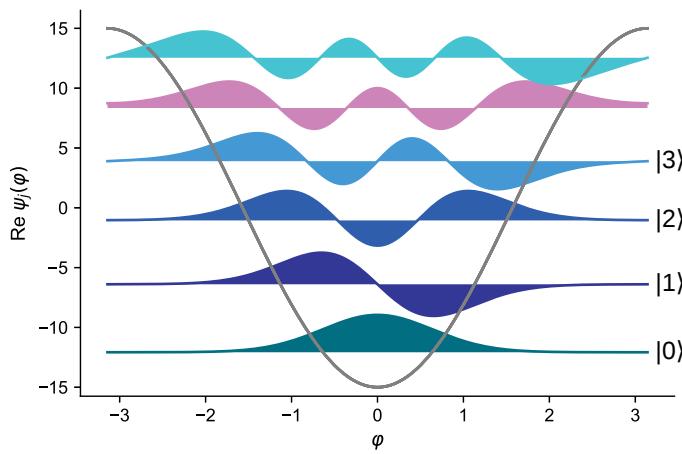


Fig. 73 Потенциал и уровни энергии в сверхпроводниковом кубите-трансмопе. Построено с помощью scqubits [GK21]. Как правило используются нижние два уровня, однако и остальные уровни можно использовать для алгоритмов

Бывают модели квантовых вычислений на непрерывных степенях свободы, например:

- оптические вычисления на непрерывных переменных;
- колебательное движение ионов в радиочастотных ловушках;
- состояния микроволнового поля в сверхпроводниковых резонаторах.

Способ выполнения операций

Модели вычислений в квантовых компьютерах можно классифицировать по признаку способа реализации операций в них. Самый простой и очевидный способ – воздействие на кубит (кубит, непрерывную переменную) при помощи короткого аналогового сигнала на резонансной частоте этой системы, причем амплитуда и фаза этого сигнала будет определять унитарную матрицу операции. Так работают:

- сверхпроводниковые кубиты-трансмоны и потоковые кубиты;
- кубиты на ионах в радиочастотных ловушках;
- кубиты на нейтральных атомах в дипольных ловушках.



Fig. 74 Схема квантового компьютера на ионах. Источник: [WBD+19]. Кубиты расположены рядом друг с другом в вакуумной камере, а управляющие сигналы представляют собой лазерные лучи, направленные на каждый ион. Для однокубитных и двухкубитных операций используют разные лазеры, для считывания – еще лазер. Рассеянное на ионе излучение последнего лазера детектируется, чтобы определить состояние кубита. Похожим образом работают и кубиты на нейтральных атомах (там есть дополнительный оптический пинцет), и сверхпроводниковые кубиты (там вместо лазеров используются микроволновые кабели).

Бывают вычислительные модели, в которых вместо временной последовательности управляющих сигналов носитель информации движется по “конвейеру”, вдоль которого с ним выполняются операции. Так работают, например, вычисления на:

- фотонах в интегральной оптике;
- фотонах в открытом пространстве.

Некоторые модели вычислений вовсе не предполагают выполнения каких-то изолированных операций. Впрочем, для таких моделей вычислений и не выполняется большинство утверждений и теоретических предсказаний, связанных с квантовыми вычислениями. Таковыми моделями является, например, **квантовый аннилип D-Wave**. Важно отметить, что классические компьютеры работают совсем иначе, чем все вышеперечисленные квантовые компьютеры. В них операции осуществляются цифровыми сигналами, то есть от непосредственного значения уровня сигнала слабо зависит то, какая операция будет выполнена, главное, чтобы сигнал вовремя превысил некоторое пороговое значение. Цифровой характер классических компьютеров имеет очень важное значение для масштабирования алгоритмов, ведь именно он обеспечивает воспроизводимость и возможность безошибочной работы квантового компьютера. К моделям квантовых вычислений, нечувствительным к аналоговым ошибкам управляющих сигналов, или с пониженной чувствительностью к ним, можно отнести:

- кубиты на майорановских фермионах;
- логические кубиты, которые получаются в результате работы алгоритмов квантовой коррекции ошибок на физических кубитах любой природы;
- сверхпроводниковые 0-пи кубиты, при условии, что гейты будут производиться специальным помехозащищенным образом, а не через промежуточные уровни.

Наборы инструкций и универсальность

Разные модели вычислений дают разные наборы инструкций для реализации квантовых алгоритмов. В зависимости от вероятности успешной реализации инструкций (квантовые компьютеры делают ошибки!) эти модели вычислений могут быть ближе, либо дальше от создания универсального масштабируемого квантового компьютера. По этому признаку:

- сверхпроводниковые кубиты;
- нейтральные атомы;
- ионы в радиочастотных ловушках

являются самыми очевидными. Набор инструкций в них содержит однокубитные и двухкубитные операции, инициализацию кубита в определенном состоянии и считывание. Чем выше точность операций – тем лучше, и все операции на сегодняшний день продемонстрированы с вероятностью положительного исхода выше 99% – этого все еще мало для многих применений, но это уже что-то.

Для вычислений на линейнооптических схемах используются источники единичных эквивалентных фотонов, либо источники запутанных фотонов, либо источники сжатого света. Вероятность успеха операции по созданию такого излучения (либо степень сжатия света) на сегодняшний день сильно далека от того, что нужно для универсальности квантовых вычислений; здесь идет речь о сжатии порядка 10 дБ (грубо переводя на язык "успешного выполнения операции" 0,9), и вероятности испускания фотона 0,9. Однако отсутствие универсальности не означает, что линейнооптические схемы должны быть для всех задач вычислительно слабее, чем другие платформы: напротив, самое сильное квантовое превосходство на сегодняшний день достигнуто именно на таких устройствах.

Причина, по которой линейнооптические системы выделяются среди других заключается в том, что для полноценного компьютера (что классического, что квантового) нужны нелинейные элементы. В классическом компьютере это транзисторы. На одних линейных индуктивностях и емкостях можно, в зависимости от способа кодирования информации, выполнять какой-то набор операций, но универсальным он не будет, потому что функция отклика линейной цепи является линейным. Алгоритмы, которые не сводятся к умножению матрицы на вектор в некотором представлении данных, на линейной системе выполнить не получится. Чтобы обойти это ограничение, линейно-оптические системы для квантовых вычислений принимают на часть своих входов, в дополнение к входным данным, сложные неклассические сигналы, а на выходе производится считывание числа фотонов, которое является квадратичным по отношению к амплитудам сигналов. Дополнение линейной оптики наборами нелинейных операций может сделать такие модели вычислений универсальными.



Fig. 75 Принципиальная схема работы оптического квантового чипа Xanadu. Взято из работы [ABBBradler+21]. а – квантовый алгоритм, выполняемый схемой: 1) сжатие света до 8 модах и создание запутанности (формирование двухмодового скатового состояния света, эквивалент однокубитных гейтов), 2) программируемые двухкубитные гейты в каждом четырехмодовом подпространстве, 3) считывание всех 8 мод в фоковском базисе; б – изображение чипа, на котором видны основные модули; в – схема управляющих сигналов и приборов; г – фотография всей системы (кроме управления детектором фотонов) в серверной стойке.

Наконец, кубиты с дискретными операциями, такие как логические кубиты и кубиты на майорановских фермионах, сами по себе не дают полный набор квантовых логических операций – в дополнение к ним нужны обычные кубиты. Впрочем, задача построения кубита на майорановских фермионах или на коррекции ошибок сама по себе настолько сложная, что необходимость связать с ним обычновенные дополнительные кубиты не кажется такой сложной.

Успехи в практической реализации

Платформы для квантовых вычислений можно отсортировать по максимальному размеру гильбертова пространства и точности операций (не только унитарных гейтов, но и считыванию и инициализации), которое удавалось реализовать экспериментально. В этом рейтинге только три верхние платформы находятся относительно близко к режиму квантового превосходства в подтвержденных экспериментальных работах. Итак, рейтинг:

- линейная оптика (фотоны) в открытом пространстве;
- сверхпроводниковые кубиты-трансмόны;
- ионы в линейных радиочастотных ловушках;
- интегральнооптические фотонны;
- нейтральные атомы в дипольных ловушках;
- спины в квантовых точках.

Другие платформы существуют на сегодняшний день в лучшем случае в виде двухкубитной системы, чаще – одного кубита, а лучше вообще в виде теоретической концепции – которая может оказаться вовсе нереализуемой на практике.

Сверхпроводниковые кубиты

Сверхпроводниковые кубиты относятся к наиболее успешным типам кубитов для квантовых вычислений и симуляции на сегодняшний день. Google впервые продемонстрировала квантовое превосходство на квантовом процессоре из 53 сверхпроводниковых кубитов-трансмόнов [AAV+19]. Кроме Google, сверхпроводниковые кубиты используют IBM (в том числе для своего открытого облачного проекта IBM Quantum Experience), Rigetti Computing и другие крупные фирмы и лаборатории по всему миру.



Fig. 76 Фотографии, изображения со сканирующего электронного микроскопа и эквивалентная электрическая схема микросхемы сверхпроводникового квантового процессора с двумя кубитами-трансмонами. Синим отмечен электрод одного из трансмонов. Микросхема была изготовлена в НОЦ ФМН в МГТУ им. Баумана, топология разработана в НИТУ МИСиС.

Квантовый процессор на сверхпроводниковых кубитах – это электрическая цепь. Этую электрическую цепь можно изобразить в схемы с квазисосредоточенными элементами.

Квантование электрических цепей

Электрические цепи, как и любые другие физические объекты, могут быть описаны с помощью законов квантовой механики [VD17]. Для этого следует начать с гамильтонового описания, например с помощью функции Лагранжа. Энергия всей физической системы (электрической цепи) должна быть записана в виде суммы энергий, накопленных ее составными частями (сосредоточенных элементов), выраженной через динамические переменные системы и их производные. Самые простые электрические элементы, такие как конденсаторы и

индукторы, запасают энергию $\langle E_C \rangle$ и $\langle E_L \rangle$ в электрическом и магнитном поле, а выражение для них можно записать либо через заряд на обкладках $\langle Q \rangle$ и поток через кольцо $\langle \Phi \rangle$, либо через напряжение $\langle V \rangle$ и ток $\langle I \rangle$, соответственно:

$$\langle E_C \rangle = \frac{Q^2}{2C} = \frac{\Phi^2}{2L} = \frac{CV^2}{2} = \frac{I^2}{2R}$$

Для третьего самого простого линейного элемента электрической цепи, резистора, так сделать не получится.

Резистор не накапливает в себе энергию, он ее диссирирует. Диссипативные системы не могут быть описаны исключительно при помощи функции Лагранжа. Это не значит, что резистор совершенно невозможно адекватно описать в гамильтоновом формализме: он может быть представлен как бесконечный резервуар энергии, так как с точки зрения остальной цепи не слишком важно, осталась ли энергия в резисторе или она была рассеяна, если она никогда больше не вернется в остальные элементы цепи. Резисторы не слишком полезны в квантовой обработке информации, так как любые потери энергии приводят к неунитарной эволюции, потере когерентности и в конечном счете – ошибкам вентиляй.

В электротехнике, как правило, используются переменные $\langle V \rangle$ и $\langle I \rangle$: напряжение и ток можно померить, правильно подключив измерительные приборы к проводникам цепи, в то время как измерения $\langle Q \rangle$ и $\langle \Phi \rangle$ затруднены, так как заряды и потоки локализованы внутри элементов. Но $\langle V \rangle$ и $\langle I \rangle$ не подходят в качестве переменных, от которых зависит функция Лагранжа, так как ни одна из них не является производной другой. Зато из закона сохранения заряда следует, что $\langle \dot{Q} \rangle = -\langle I \rangle$, а закон индукции Фарадея говорит, что $\langle \dot{\Phi} \rangle = \langle V \rangle$, а значит и $\langle \Phi \rangle$, и $\langle Q \rangle$ могут быть использованы в качестве динамических переменных системы.



Fig. 77 Электрические схемы одного LC-контура и двух емкостно связанных LC-контуров.

В случае одной степени свободы определения для величин $\langle Q \rangle$, $\langle \Phi \rangle$, $\langle V \rangle$ и $\langle I \rangle$ однозначны. В более сложных случаях в классической электротехнике принято использовать метод контурных токов, либо узловых напряжений. Рассмотрим метод узловых напряжений. Для лагранжевого описания введем узловой поток $\langle \Phi_i(t) \rangle$ как первообразную напряжения на узле с номером i $\langle V_i(t) \rangle$:

$$\langle \Phi_i(t) \rangle = \int \limits_{t_0}^t V_i(t) dt$$

Функция Лагранжа LC-контура имеет вид

$$L(\langle \Phi_1 \rangle, \dot{\langle \Phi_1 \rangle}) = \frac{C \dot{\langle \Phi_1 \rangle}^2}{2} - \frac{\langle \Phi_1 \rangle^2}{2L}$$

Канонически сопряженный импульс $\langle Q_i \rangle$ к обобщенной переменной $\langle \Phi_i \rangle$ можно найти по формуле

$$\langle Q_i \rangle = \frac{\partial L}{\partial \dot{\langle \Phi_i \rangle}} = C \dot{\langle \Phi_i \rangle}$$

Физический смысл этого импульса – сумма зарядов на конденсаторах, примыкающих к узлу i , или узловой заряд.

Функция Гамильтона $\langle H(\langle \Phi_i \rangle, \langle Q_i \rangle) \rangle$ представляет собой преобразование Лежандра от функции Лагранжа по парам $\langle (\dot{\langle \Phi_i \rangle}, \langle Q_i \rangle) \rangle$, то есть

$$\langle H(\langle \Phi_i \rangle, \langle Q_i \rangle) \rangle = \dot{\langle \Phi_i \rangle} \dot{\langle Q_i \rangle} - L = \frac{\langle Q_i \rangle^2}{2C} + \frac{\langle \Phi_i \rangle^2}{2L}$$

Оператор Гамильтона может быть получен из функции Гамильтона заменой переменных $\langle \Phi_1, Q_1 \rangle$ на соответствующие операторы:

$$H = \frac{\langle \hat{Q}_1 \rangle^2}{2C} + \frac{\langle \hat{\Phi}_1 \rangle^2}{2L}$$

Здесь $\langle \hat{Q}_1, \hat{\Phi}_1 \rangle = i\hbar$ в силу того, что $\langle \Phi_1 \rangle$ и $\langle Q_1 \rangle$ – каноническая переменная и сопряженный к ней импульс. Волновая функция системы может быть записана в потоковом $\langle \psi(\langle \Phi_1 \rangle) \rangle$ или зарядовом $\langle \psi(Q_1) \rangle$ представлении. Переход от одного представления к другому, как в случае обыкновенных координат и импульсов, может быть произведен с помощью преобразования Фурье

$$\langle \psi(\langle \Phi_1 \rangle) \rangle = \frac{1}{\sqrt{2\pi}} \int e^{i\langle \Phi_1 \rangle p} \langle \psi(Q_1) \rangle dQ_1$$

Как и в случае механических осцилляторов, для LC-осцилляторов очень полезен формализм операторов вторичного квантования $\langle \hat{a} \rangle$ и $\langle \hat{a}^\dagger \rangle$, которые можно определить как

$$\langle \hat{a} \rangle = \frac{\langle \hat{Q}_1 \rangle}{\sqrt{2\pi}} \langle \hat{\Phi}_1 \rangle / \langle \hat{Q}_1 \rangle$$

где $\langle Z \rangle = \sqrt{L/C}$ – импеданс осциллятора. В этом формализме гамильтониан системы принимает вид

$$\langle \hat{H} \rangle = \hbar \omega_1 \langle \hat{a}^\dagger \hat{a} \rangle$$

где $\langle \omega_1 \rangle = (L/C)^{1/2}$ – частота осциллятора.

Линейные электрические элементы, такие как емкости и индуктивности, очень полезны при создании электронных приборов. Но из линейных элементов можно создать только линейную электрическую цепь. Это означает, что любые выходные сигналы $\langle V^{out} \rangle$ выражаются через входные сигналы $\langle V^{in} \rangle$ с помощью линейного преобразования

$$\langle V^{out} \rangle_i(\omega) = \sum \limits_j S_{ij}(\omega) \langle V^{in} \rangle_j(\omega)$$

Это выражение остается справедливым и для квантового случая, где величина $\langle V^{out} \rangle_i(\omega)$ является оператором. $\langle S_{ij}(\omega) \rangle$ - это передаточная функция, или S -параметр системы. Для линейной системы S -параметр зависит лишь от частоты сигнала, но не зависит от состояния системы. По существу линейные системы не обладают состоянием, т.е. не могут играть роль памяти. В частности, такие системы не могут эффективно симулировать машину Тьюринга и тем более квантовую машину Тьюринга. Они могут реализовывать некоторое линейное преобразование над входными данными, однако это линейное преобразование жестко задано топологией и номиналами емкостей и индуктивностей цепи.

Для того, чтобы создать архитектуру компьютера, в которой есть память или управление, нужны нелинейные элементы. В классической электронике в качестве нелинейного элемента чаще всего используют транзистор. Наиболее популярный в интегральных микросхемах - это полевой транзистор. Полевой транзистор имеет три контакта, причем напряжение между затвором и стоком управляет сопротивлением между стоком и истоком. Для квантовой обработки информации управление сопротивлением – не самый практичный вариант, в связи с тем, что протекание тока через конечные сопротивления приводит к ошибкам квантовых вентилей.

В качестве альтернативы полупроводниковым нелинейным элементам вроде транзистора и диода можно использовать сверхпроводниковые нелинейные элементы. Одним из таких элементов является джозефсоновский контакт сверхпроводник-изолятор-сверхпроводник (Superconductor-insulator-superconductor, SIS).

Для гамильтонового описания джозефсоновского контакта как элемента электрической цепи используется джозефсоновская энергия, которая определяется соотношением

$$E_J = \frac{I_c \Phi_0}{2\pi} (1 - \cos \phi)$$

где I_c – критический ток контакта, $\Phi_0 = h/2e$ – квант магнитного потока в сверхпроводнике, $\phi = 2\pi \Phi / \Phi_0$ – обезразмеренная на квант потока разность узловых потоках на берегах контакта. Эта величина одновременно является фазой конденсата куперовских пар в сверхпроводнике. Для понимания принципов работы сверхпроводниковых кубитов совершенно необходимо понимать физику самой сверхпроводимости, что такое куперовская пара, что подразумевается под словом конденсат и что значит его фаза. С точки зрения схемотехники кубитов, все знания о физике сверхпроводников могут быть сведены к определению фазы и соотношением для энергии контакта в зависимости от фазы. Эти соотношения однозначно определяют поведение потоков и зарядов в цепи. Мы будем пользоваться фазой вместо узлового потока исключительно из-за того, что эта величина является безразмерной, и это позволит упростить все соотношения. Также заряд будет измеряться в единицах заряда куперовской пары, то есть $(2e)$.

Дифференцируя соотношение для джозефсоновской энергии по потоку, можно получить выражение для индуктивности джозефсоновского контакта

$$L_J = \frac{1}{2} \frac{\partial E_J}{\partial \Phi} = \frac{I_c \Phi_0^2}{2\pi} (2\pi)^2 \cos \phi = \frac{I_c \Phi_0^2}{2\pi} \cos \phi$$

Трансмон



Fig. 78 Электрическая схема трансмона. I_c – критический ток контакта, C_s – шунтирующая емкость, C_j – емкость контакта

Самый простой, и в то же время один самых успешных видов сверхпроводниковых кубитов – это трансмон [KYG+07]. Трансмон представляет собой ангармонический осциллятор на основе наноразмерного джозефсоновского контакта, шунтированного емкостью. Такая цепь хорошо известна в сверхпроводниковой электронике, а ее резонансная частота называется плазменной частотой:

$$\omega_p = \sqrt{L_J(C_j + C_s)}$$

В квантовом случае, свойства этого осциллятора могут зависеть от постоянного напряжения $\langle V_g \rangle$. Наступление квантового режима определяется произведением критического тока и емкости; фактически, в зависимости от этого произведения, такая схема может быть классическим осциллятором, трансмоном или зарядовым кубитом.

Гамильтониан трансмона (а также зарядового кубита) имеет вид

$$\langle \hat{H} \rangle = \frac{1}{2}(n\langle g \rangle)^2(2C_J + C_S) + \frac{1}{2}C_J\langle \Phi_0 \rangle^2(1 - \cos\langle \Phi \rangle),$$

где $\langle \hat{n} \rangle = \langle Q \rangle / 2e$, $\langle \hat{\Phi} \rangle = \frac{1}{2}\langle \hbar \rangle \langle \Phi \rangle$ – безразмерные операторы числа и фазы

куперовских пар на электроде, а $\langle n \langle g \rangle \rangle$ – средний наведенный на острове внешними источниками

напряжения заряд. Гамильтониан можно упростить, заменив произведения размерных константы на параметры

$$E_C = \frac{e^2 n \langle g \rangle^2 (C_J + C_S)}{2e \langle \Phi_0 \rangle^2 \langle \Phi \rangle},$$

которые называют емкостной и джозефсоновской энергией трансмона, соответственно. В этих обозначениях

гамильтониан принимает вид

$$\langle \hat{H} \rangle = 4E_C(\langle \hat{n} \rangle - n_g)^2 + E_J(1 - \cos\langle \hat{\Phi} \rangle).$$

Уровни энергии стационарных состояний в такой системе будут зависеть от приложенного напряжения (вернее, от наведенного внешним напряжением $\langle V_g \rangle$ зарядом $\langle n_g \rangle$). Однако эта зависимость ослабевает по мере того, как увеличивается соотношение (E_J/E_C) .

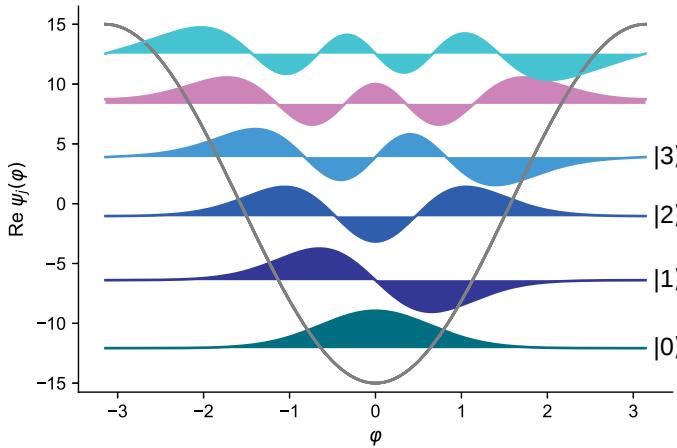


Fig. 79 Уровни энергии в зарядовом кубите и в трансмоне

Трансмоном называется режим этой электрической цепи, когда $(E_J/E_C \approx 50)$; при таком соотношении между константами зависимость от заряда практически отсутствует. Это очень хорошо, потому что наведенный заряд очень сложно контролировать на практике. Космические частицы и события радиоактивного распада в подложке приводят к запиранию зарядов в случайных местах в подложке, из-за чего зарядово-чувствительные кубиты могут менять частоту перехода между уровнями. Уровни энергии в трансмоне не чувствительны к таким воздействиям, что значительно улучшает их времена когерентности по сравнению с зарядовыми кубитами.

Обратной стороной такой схемы является невысокий ангармонизм. Ангармонизмом (δ) называется разность частот между переходами $(|0\rangle \leftrightarrow |1\rangle)$ и $(|1\rangle \leftrightarrow |2\rangle)$. У трансмона $(\delta = -E_C)$. По существу, трансмон является слабо ангармоничным осциллятором. Однако расстояния между частотами переходов в трансмоне достаточно большие, чтобы можно было попасть в один переход и при этом не попасть в соседний.

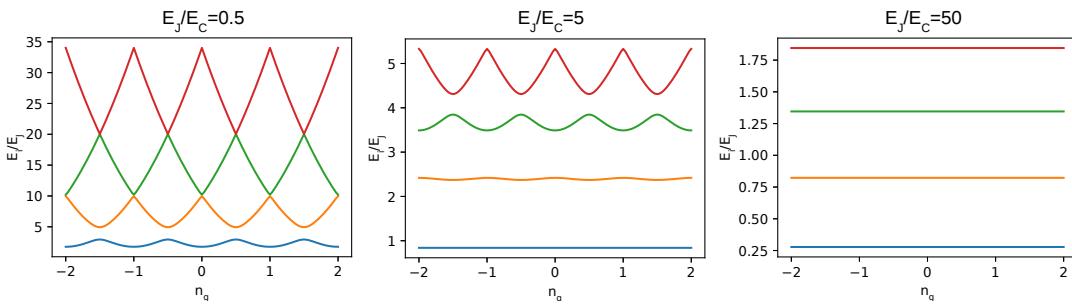


Fig. 80 Уровни энергии и волновые функции стационарных состояний трансмона с $(E_J/E_C=50)$

Считывание состояния кубита в сверхпроводящих цепях, как правило, осуществляется методом дисперсионного считывания [WSB+05]. В основе дисперсионного считывания состояния сверхпроводниковых кубитов стоит архитектура квантовой электродинамики в электрической цепи, в которой сверхпроводниковый кубит связан с микроволновым резонатором. Концепция квантовой электродинамики в электрической цепи возникла по аналогии с квантовой электродинамикой в полости, где уединенный атом помещается в оптическую полость, которая выступает в роли резонатора. Взаимодействие такого атома с излучением сильно отличается от того, что происходит с таким атомом, помещенным в вакуум: из-за отражений от стенок полости, излучение не может эффективно покинуть атом. Это увеличивает время жизни атома, если резонансная частота полости находится вдали от спектральной линии кубита. Напротив, если атом находится в резонансе с полостью, то это стимулирует излучение.

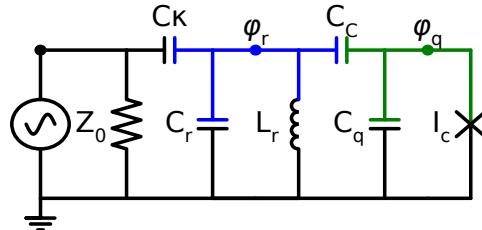


Fig. 81 Эквивалентная электрическая схема трансмона, связанного с резонатором. Синий — электрод резонатора, зеленый — электрод трансмона

Для многокубитных схем, как правило, в качестве микроволнового резонатора используется участок копланарной передающей линии на микросхеме. Микроволновый резонатор, в свою очередь, подключается к микроволновой линии.

Пропускание микроволновых сигналов через линию приводит к изменению квантового состояния электромагнитного излучения в резонаторе; это излучение взаимодействует с кубитом. Преимуществом данной схемы является то, что кубит может быть очень сильно связан с резонатором, но при этом сохраняет когерентность. Если пренебречь влиянием потерь в управляющей линии, то система описывается гамильтонианом вида

$$[\hat{H}] = \left(\omega_q + \frac{1}{2} \delta_q \right) \hat{a}^\dagger \hat{a} - \left(\omega_r + \frac{1}{2} \delta_r \right) \hat{b}^\dagger \hat{b} + i \chi \left(\hat{a}^\dagger \hat{b} + \frac{1}{2} \right)$$

где ω_q, ω_r — частоты кубита и резонатора, соответственно; δ_q, δ_r — ангармонизмы мод трансмона и резонатора; $i\chi$ — дисперсионный сдвиг, или кросс-керровская нелинейность между трансмоном и резонатором, \hat{a}, \hat{b} — одночастичные операторы уничтожения и рождения, действующие на моду трансмона, $\hat{a}^\dagger, \hat{b}^\dagger$ — одночастичные операторы рождения и уничтожения, действующие на моду резонатора, $\Omega(t)$ — амплитуда возбуждения со стороны микроволновой линии.

Если емкость C_C между трансмоном и резонатором маленькая, а их частоты сильно отличаются, то взаимодействие между ними будет нерезонансным, и тогда ангармонизмом резонатора можно пренебречь. А вот наличие дисперсионного сдвига между модой кубита и резонатора приводит к тому, что частота эффективная резонатора, то есть сумма всех чисел перед множителем $\hat{b}^\dagger \hat{b}$, зависит от населенности трансмона. В таком случае, измеряя отклик от резонатора на микроволновый импульс вблизи его резонансной частоты, мы можем получить информацию о состоянии кубита.

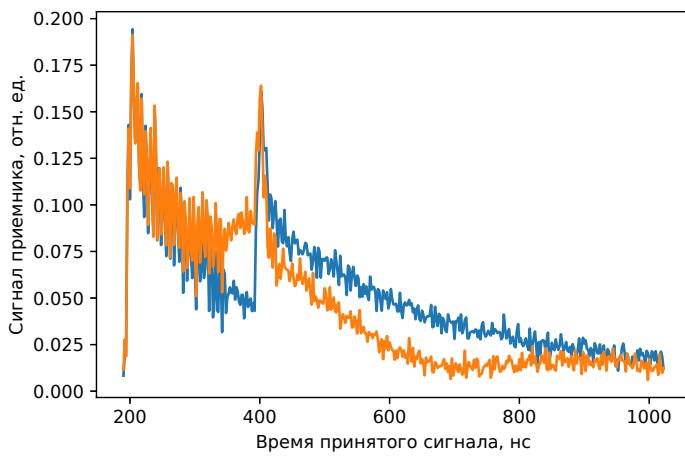


Fig. 82 Микроволновый прямоугольный импульс, отраженный от резонатора, при различных состояниях трансмона. Усреднено по 10000 повторениям.

Микроволновые однокубитные операции

Логические операции над кубитами выполняют микроволновыми сигналами на резонансной частоте кубита.

Поведение трансмона по отношению к внешнему излучению можно представить в виде гамильтониана

$$\hat{V} = V_{\text{ex1}}(t) \frac{C_g + C_{qs}}{C_g + C_{qs}} \hat{Q}$$

для схемы с возбуждением напряжением $V_{\text{ex1}}(t)$, подаваемым через маленький конденсатор. Пусть

$$V_{\text{ex1}}(t) = V_I(t) \cos(\omega_q t + \phi(t)) + V_Q(t) \sin(\omega_q t + \phi(t)),$$

где $(V_I(t), V_Q(t))$ – медленно изменяющиеся огибающие напряжения, а $(\phi(t))$ – медленно изменяющаяся фаза. “Медленно изменяющаяся” здесь означает, что за переходы между уровнями будет отвечать только быстро осциллирующий множитель. Обозначим

$$\Omega(t) = I(t) - iQ(t) = \frac{C_g + C_{qs}}{\sqrt{\hbar/2Z}} (V_I - iV_Q)$$

Полный гамильтониан трансмона с возбуждающим излучением будет иметь вид

$$\begin{aligned} \hat{H} &= \frac{1}{\hbar} (\omega_q \hat{a}^\dagger \hat{a} + \frac{1}{2} \delta \omega_q (\hat{a}^\dagger \hat{a} - 1) + \Omega(t) e^{i(\omega_q t + \phi(t))}) \\ &+ \Omega^*(t) e^{-i(\omega_q t + \phi(t))} (\hat{a}^\dagger + \hat{a}) \end{aligned}$$

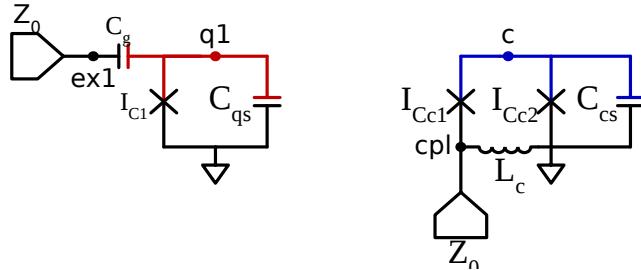


Fig. 83 Трансмон с линиями для микроволновых однокубитных операций. Слева – емкостно связанный линии (через конденсатор), справа – индуктивно связанный линии (через индуктивность).

Если трансмон использовать как кубит (то есть не возбуждать никаких уровней, кроме основного и первого возбужденного), можно перейти от операторов повышения и понижения к операторам Паули с помощью соотношений

$$\begin{aligned} \hat{a}^\dagger \hat{a} &= -\frac{1}{2} (\hat{\sigma}_z + \hat{a}^\dagger + \hat{a}) \\ \hat{a}^\dagger &= \hat{\sigma}_x \end{aligned}$$

Тогда мы получим гамильтониан вида

$$\hat{H} = \frac{1}{\hbar} (\Omega(t) e^{i(\omega_q t + \phi(t))} + \Omega^*(t) e^{-i(\omega_q t + \phi(t))}) \hat{\sigma}_x$$

Такое возбуждение отвечает модели Раби для двухуровневой системы. Модель Раби хорошо известна в квантовой оптике, так как она описывает когерентное взаимодействие одного атома с лазерным излучением.

Здесь мы остановимся поподробнее, так как модель Раби имеет центральное значение для однокубитных операций на очень широком классе различных кубитов.

Пусть $\langle |\psi(t)\rangle = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}$ – решение уравнения Шредингера. Для того, чтобы решить уравнение Шредингера, которое является системой двух обыкновенных линейных дифференциальных уравнений относительно $|c_0(t)\rangle$ и $|c_1(t)\rangle$, нужно сделать замену

$$\begin{aligned} & \langle |c_0(t) = d_0(t)e^{-i\omega_q t/2}, |c_1(t) = d_1(t)e^{+i\omega_q t/2}. \end{aligned}$$

Эта замена называется переходом во вращающуюся систему отсчета. В данном случае, система отсчета вращается вместе с быстро осциллирующей частью микроволнового сигнала. Эта замена может быть записана в виде произведения матрицы на вектор:

$$\begin{aligned} & \begin{aligned} & \langle |c_0(t) \&= d_0(t) e^{-i\omega_q t/2} \\ & \& d_1(t) = d_1(t) e^{+i\omega_q t/2}. \end{aligned} \end{aligned}$$

Матрица этого линейного преобразования $\langle |\hat{U}| = e^{\langle i(\omega_q t + \phi(t)/2)} \hat{|\sigma_z|}$) является унитарной – ее собственные числа по модулю равны единице. Если подставить эту замену в уравнение Шредингера

$$\langle i\hbar \partial_t |\psi\rangle = \hat{H} |\psi\rangle,$$

то получим систему

$$\begin{aligned} & \begin{aligned} & \langle |d_0 \&= e^{\langle i(\omega_q t + \phi(t)/2)} + \frac{\dot{\omega}_q + \dot{\phi}}{2} d_0 \\ & \& d_1 \&= -\frac{\dot{\omega}_q}{2} d_0 e^{\langle i(\omega_q t + \phi(t)/2)} + d_1 e^{\langle i(\omega_q t + \phi(t)/2)} \\ & \& \left. \langle |d_0 \&= -\frac{\dot{\omega}_q}{2} e^{\langle i(\omega_q t + \phi(t)/2)} - \frac{\dot{\omega}_q + \dot{\phi}}{2} d_1 \\ & \& d_1 \&= \frac{\dot{\omega}_q}{2} e^{\langle i(\omega_q t + \phi(t)/2)} + d_0 e^{\langle i(\omega_q t + \phi(t)/2)}. \end{aligned} \end{aligned}$$

Нехитрыми алгебраическими преобразованиями можно упростить гамильтониан до вида

$$\begin{aligned} & \begin{aligned} & \langle |d_0 \&= \frac{\dot{\omega}_q}{2} d_0 + \langle |d_1 \&= \frac{\dot{\omega}_q}{2} d_1 \\ & \& \langle |d_0 \&= -\frac{\dot{\omega}_q}{2} d_1 + \langle |d_1 \&= \frac{\dot{\omega}_q}{2} d_0. \end{aligned} \end{aligned}$$

Далее используют “приближение вращающейся волны”, смысл которого в том, чтобы отбросить быстро осциллирующие слагаемые в гамильтониане с множителями $(e^{-2\omega_d t})$ и $(e^{2\omega_d t})$. При не очень больших амплитудах возбуждающих сигналов $(|\Omega(t)| \ll \omega)$, эти слагаемые слишком быстро меняются, чтобы оказать существенное влияние на более медленное изменение $(d_0(t))$ и $(d_1(t))$.

Получается система уравнений уже с постоянными коэффициентами

$$\begin{aligned} & \begin{aligned} & \langle |d_0 \&= \frac{\dot{\omega}_q}{2} d_0 + \langle |d_1 \&= -\frac{\dot{\omega}_q}{2} d_1 \\ & \& \langle |d_0 \&= -\frac{\dot{\omega}_q}{2} d_0, \langle |d_1 \&= \frac{\dot{\omega}_q}{2} d_1. \end{aligned} \end{aligned}$$

которую можно записать через операторы Паули в виде

$$\langle |i\tilde{|\psi\rangle} = \langle |I(t)\hat{|\sigma_x} + Q(t)\hat{|\sigma_y} - \frac{\dot{\omega}_q}{2}\hat{|\sigma_z}| \tilde{|\psi\rangle}, \rangle$$

где $\langle |\tilde{|\psi\rangle} = \begin{pmatrix} d_0(t) \\ d_1(t) \end{pmatrix}$.

Такой гамильтониан отвечает вращению вектора состояния кубита в представлении сферы Блоха вокруг оси $\langle |\vec{\Omega}(t)| = \langle |I(t), Q(t), \dot{\phi}(t)/2\rangle$ с угловой скоростью $\langle |\vec{\omega}(t)|$.

Наиболее часто используются микроволновые импульсы, отвечающие вращениям на углы $\langle |\pi/2\rangle$ и $\langle |\pi\rangle$ вокруг осей $\langle |x\rangle$ или $\langle |y\rangle$, однако реально любая однокубитная операция может быть реализована правильным подбором управляющих сигналов $\langle |I(t)|$, $\langle |Q(t)|$, или $\langle |\phi(t)|$.

Декогеренция в сверхпроводниковых кубитах

В отличие от других типов кубитов, сверхпроводниковые кубиты имеют сравнительно низкие времена когерентности. Это означает, что информация, хранящаяся в сверхпроводниковом кубите, может испортиться. Для сверхпроводниковых кубитов декогеренцию можно понимать как неконтролируемое воздействие на кубит со стороны флюктуаций напряжения или тока. Природа флюктуаций может быть очень разной. Наиболее проблематичными являются флюктуации микроскопических зарядов и магнитных моментов внутри самой микросхемы.

Как правило различают шумы и наводки, которые могут привести самопроизвольному возбуждению или релаксации кубита, а также к переходу в недопустимые состояния типа уровней $\langle |2\rangle$ и выше, и шумы, которые влияют на частоту переходов между уровнями кубита.

В самой простой, марковской модели шума, события декогеренции независимы во времени и пространстве. Флюктуационный сигнал $\langle |S_f(t)|$, связывающийся с трансмоном посредством члена гамильтониана типа $\langle |\hat{V}| = S_f(t) \hat{|\sigma_a}^\dagger \hat{|\sigma_a}|$, будет случайным образом менять частоту кубита. Это приведет к набегу фазы

$$\langle |\Delta\phi(t) = \int \limits_{0}^{t} S_f(t') dt' \rangle$$

на кубите. Если разложить флюктуационный сигнал по спектральным компонентам (то есть представить в виде Фурье-компонент), то получим

$$\langle \Delta \phi(t) = \int_{-\infty}^{+\infty} \frac{S_f(\omega)}{\omega} \cos\left(\frac{\omega t}{2}\right) d\omega \rangle$$

Таким образом, наиболее существенным с точки зрения случайного набега фазы (то есть дефазировки) является низкочастотный шум. К таким флюктуационным сигналам может относиться, например, тепловая населенность считающего резонатора.

Другой тип флюктуационного сигнала связывается с трансмоном посредством дополнительного члена в гамильтониане, имеющего вид $\langle \hat{V} = i \Delta \Omega \downarrow \hat{a} - \hat{a}^\dagger \rangle$. Такой флюктуационный сигнал приводит к самопроизвольному возбуждению и релаксации кубита под действием внешней силы. Наиболее существенными являются шумы, частота которых отвечает частоте кубита.

Помимо внешних воздействий, которые можно представить как случайное слагаемое в гамильтониане, существуют также и воздействия, которые нельзя представить в таком виде, например обыкновенную релаксацию кубита без сопутствующего теплового возбуждения. Такой канал декогеренции отвечает случайному действию оператора уничтожения $\langle \hat{a} \rangle$. Таковым является, например, релаксация трансмона в управляющую линию, через которую выполняются однокубитные операции.

В общем случае, для всех типов марковского шума можно ввести оператор коллапса $\langle \hat{c} \rangle$ и константу декогеренции $\langle \gamma \rangle$. В случае релаксации, $\langle \hat{c} \rangle = \hat{a}, \gamma = 1/T_1$. Для дефазировки $\langle \hat{c} \rangle = \langle \hat{\sigma}_z \rangle, \gamma = 1/2T_{\phi}$.

Для того, чтобы измерить эффекты декогеренции на сверхпроводниковые кубиты, используют два основных типа эксперимента. Первый типа эксперимента заключается в возбуждении кубита и считывании его состояния спустя время $\langle t \rangle$. Эксперимент повторяется много раз с различным временем задержки $\langle t \rangle$; результат можно зафитовать экспонентой. Время затухания экспоненты в $\langle e \rangle$ будет временем релаксации $\langle T_1 \rangle$.

Такое измерение нечувствительно к дефазировке кубита. Чтобы измерить влияние дефазировки, кубит помещают в состояние суперпозиции при помощи вращения вокруг оси $\langle x \rangle$ на угол $\langle \pi/2 \rangle$, ждут время $\langle t \rangle$, делают вращение вокруг оси $\langle z \rangle$ на угол, пропорциональный длительности паузы $\langle t \rangle$, а затем еще поворачивают на угол $\langle \pi/2 \rangle$ вокруг оси $\langle x \rangle$. В случае идеальной двухкубитной системы зависимость результата измерений от длительности паузы $\langle t \rangle$ будет синусоидальной; при наличии декогеренции в марковской модели осцилляции будут затухать экспоненциально, причем константа затухания $\langle T_2 \rangle$ может быть выражена как $\langle \frac{1}{T_2} \rangle = \frac{1}{2T_1} + \frac{1}{T_\phi}$.

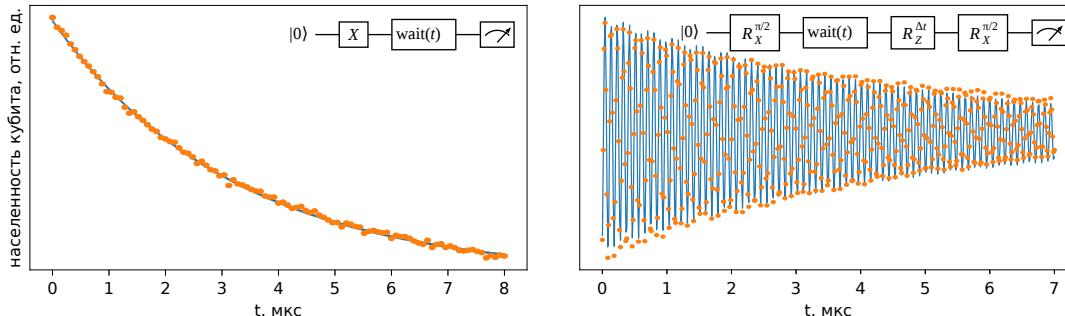


Fig. 84 Эксперименты по измерению времени когерентности кубита

Процессы декогеренции ограничивают длину алгоритмов, которые могут быть выполнены на сверхпроводниковом квантовом процессоре. Если хотя бы одно событие декогеренции произойдет в течение алгоритма, то результат выполнения может оказаться неправильным. Важно, что считывание в квантовом компьютере имеет принципиально дискретный характер: из-за этого в принципе возможно говорить о дискретных ошибках, а не о непрерывных ошибках, как в аналоговых компьютерах. Кроме того, как и в обычных цифровых компьютерах, в квантовых компьютерах возможна коррекция ошибок. Для того, чтобы измерить вероятность ошибки напрямую, используют рандомизированное тестирование (randomized benchmarking) квантовых вентилей.

Идея рандомизированного тестирования заключается в том, чтобы выполнять случайную последовательность операций, а зачем сравнивать результат ее выполнения с теоретическим. Существует ряд различных вариаций рандомизированного тестирования, которые отличаются тем, из какой совокупности берутся случайные последовательности. В самом простом случае используют вентили из группы Клиффорда – это такое подмножество логических операций, которые не образуют универсальный набор, но зато алгоритмы целиком состоящие только из таких вентилей можно быстро просимулировать на классическом компьютере даже при большом количестве кубитов. Кроме того, часто используют только такие последовательности вентилей, в конце

которых система оказывается не в состоянии суперпозиции, а в определенном состоянии, например в состоянии $|0\dots0\rangle$. Ошибки приводят к тому, что вероятность $\langle F_n | \rangle$ получить состояние $|0\dots0\rangle$ экспоненциально убывает с увеличением длины последовательности вентиляй N :

$|\langle F_n | \rangle| \sim e^{-F_1 n}$

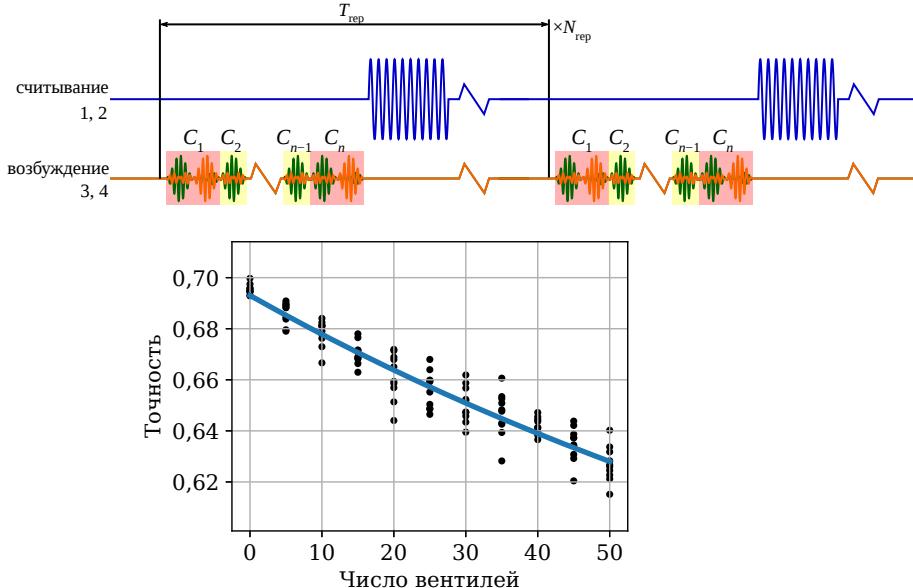


Fig. 85 Схема микроволновых импульсов и пример результатов рандомизированного тестирования вентиляй из группы Клиффорда. Каналы 1 и 2 генератора сигналов произвольной формы подключены к линии считывания, каналы 3 и 4 – к линии управления кубитом. Измерена населенность кубита после случайных последовательностей вентиляй разных длин; константа экспоненциального затухания населенности дает точность в расчете на один вентиль ($F_1 = 0.993$).

Двухкубитные вентили

Для трансмонов существует целый ряд способов реализации двухкубитных операций. Как правило, для этого трансмоны связывают емкостным способом, напрямую или через элемент связи. Мы будем рассматривать двухкубитную операцию iSWAP, реализуемую при помощи параметрического изменения эффективной константы связи между кубитами.

Так же, как в случае кубита, связанного с резонатором, гамильтониан двух связанных трансмонов описывается в первом приближении соотношением

$$\begin{aligned} \hat{H}/\hbar = & \left(\omega_1 + \frac{\delta_q}{2} \right) \hat{a}^\dagger \hat{a} + \left(\omega_2 + \frac{\delta_r}{2} \right) \hat{b}^\dagger \hat{b} + \chi \left(\hat{a}^\dagger \hat{b} + \hat{a} \hat{b}^\dagger \right) \end{aligned}$$

где \hat{a} , \hat{b} – операторы уничтожения и рождения одной степени свободы (трансмона), а χ – операторы другой степени свободы (трансмона).

Принцип работы элемента связи в том, что вместо одного джозефсоновского контакта, в нем есть два джозефсоновских контакта, соединенных параллельно. Такая цепь называется ПТ-СКВИД, или просто СКВИД (Superconducting Quantum Interference Device). Если у нас есть два джозефсоновских контакта в кольце, то такая структура будет вести себя по отношению к внешней цепи так же, как один контакт, но с управляемым критическим током. Управление критическим током можно осуществлять при помощи магнитного потока Φ_x , который проходит через кольцо СКВИДа. Соотношение для критического тока выглядит следующим образом:

$$I_{\text{SQUID}}(\Phi_x) = \sqrt{I_{C1}^2 + I_{C2}^2} \cos^2(\pi \Phi_x / \Phi_0) + (I_{C1} - I_{C2})^2 \sin^2(\pi \Phi_x / \Phi_0)$$

Таким образом, элемент связи представляет собой слабо нелинейный LC-контура с перестраиваемой индуктивностью, а значит и частотой. В зависимости от его частоты будет менять эффективная константа связи между кубитами χ , а также – в меньшей мере – и остальные параметры системы.

Когда нет двухкубитной операции, желательно, чтобы кубиты не взаимодействовали, то есть $\chi=0$. Все константы этого гамильтониана зависят от внешнего параметра – потока в СКВИДЕ элемента связи Φ_x . При быстром изменении этого параметра появляется дополнительное (диабатическое) слагаемое

$$\Omega \hat{V} = \hbar \partial_\Phi \hat{U} \frac{\partial \hat{U}^\dagger}{\partial \Phi} \frac{\partial \Phi}{\partial t} \sim \Omega \left(\hat{a}^\dagger \hat{a} + \hat{b}^\dagger \hat{b} \right)$$

Мы не будем здесь подробно рассматривать природу этого слагаемого; подробнее об этом можно почитать, например, в работе [RGM+17]. Ограничимся тем, что это взаимодействие является поперечным, то есть $\langle xx \rangle$ -взаимодействием. Если модулировать поток в СКВИде в резонансе с разностью частот между кубитами ($\omega_m = \omega_1 - \omega_2$), где ω_n – номер гармоники), то возбуждение с одного кубита будет переходить на другой, и обратно.

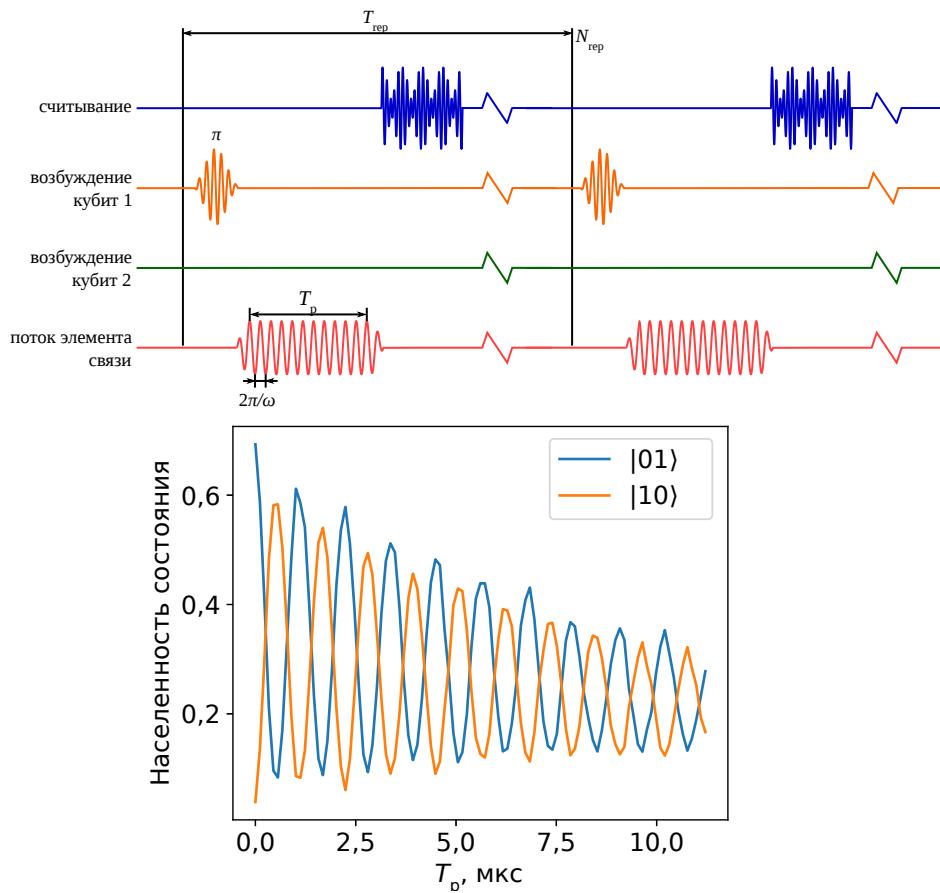


Fig. 86 Последовательность сигналов, которую нужно подать, чтобы наблюдать осцилляции населенности между кубитами, и результат из наблюдения. iSWAP получается при длительности вентиля $(T_p=500\text{--}1\text{ ns})$

Унитарная матрица этого вентиля имеет вид

$$\begin{aligned} \hat{U}(t) = & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos\Omega_g t & i \sin\Omega_g t & 0 & 0 & i \sin\Omega_g t \\ 0 & i \sin\Omega_g t & \cos\Omega_g t & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

Помимо вентиля iSWAP, на трансмонах делают CPHASE, который физически можно реализовать импульсным увеличением константы $|\chi|$, или, например, как IBM Quantum Experience – при помощи кросс-резонансных вентилей.

Что почитать для дальнейшего изучения вопроса?

- [Introduction to Quantum Computing and Quantum Hardware](#)
- [A Quantum Engineer's Guide to Superconducting Qubits \[KKY+19\]](#)

О блоке “Квантовое машинное обучение”

Этот блок включает в себя:

- общий рассказ о видах квантового машинного обучения;

Продвинутые темы блока дополнительно рассказывают:

- об алгоритме HHL;
- о квантовом алгоритме ближайшего соседа;

О квантовом машинном обучении

Квантовое машинное обучение – это на самом деле огромная область, и наш курс охватывает лишь небольшую ее часть. В этой обзорной лекции мы попробуем взглянуть на картину в целом и посмотреть, какие есть направления в этой интересной и динамично развивающейся области.

Note

Эта статья является исключительно обзорной – в ней будет много красивых картинок, но мы не будем вдаваться в формулы, которые могут эти картинки объяснить. Цель лекции – познакомить читателя с тем, как машинное обучение может быть применено в разных задачах, связанных с квантовыми вычислениями и квантовой физикой.

Виды квантового машинного обучения

Как мы помним из лекции по машинному обучению, в таких задачах у нас есть две важных составляющих – данные и алгоритм. В случае квантового машинного обучения у нас появляется некоторая “вариативность”:

- квантовые данные
- квантовый алгоритм
- классические данные
- классический алгоритм

На базе этого можно построить классификацию машинного обучения.

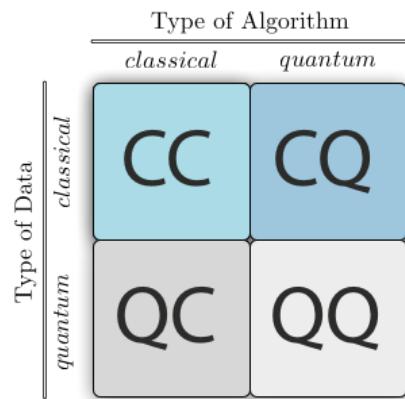


Fig. 87 Виды машинного обучения в зависимости от типа данных и алгоритма.

Когда мы имеем классические данные и классический алгоритм, то это как раз обычное машинное обучение, рассмотренное во вводной лекции курса. А вот остальные виды мы рассмотрим далее.

Квантовые данные и классический алгоритм

Классическое машинное обучение применительно к квантовым данным – это очень активно развивающаяся область квантовой физики и квантовой химии. Главный вопрос в данном случае – что именно мы понимаем под “квантовыми данными”. Но мы будем придерживаться максимально широкой трактовки этого понятия. Далее рассмотрим примеры возможного применения ML в задачах квантовой физики и химии.

NQS

Neural Quantum States это очень перспективное улучшение квантового Монте-Карло – популярного метода приближенного решения задач квантовой физики. Мы будем подробно рассматривать проблему нахождения энергетических спектров квантомеханических систем, а также моделирования квантовой динамики в блоке о проблемах квантового машинного обучения. Пока нам достаточно знания того, что эти задачи являются вычислительно трудными и их прямое численное решение очень быстро становится невозможным по мере увеличения размера системы и числа взаимодействующих частиц. В подходе NQS, впервые опубликованном в журнале *Science* [CT17], предложено моделировать волновую функцию $\langle \Psi | \Psi \rangle$ физической системы при помощи глубокой нейронной сети. Это дает ряд преимуществ над обычным квантовым Монте-Карло в плане точности моделирования, а также позволяет явно симулировать *time-dependent* уравнение Шредингера.

Относительно другого популярного метода, который называется *Matrix Product State NQS* дает лучшую масштабируемость и меньшую аппроксимацию сложность, которая оказывается линейной по числу частиц в физической системе.



Fig. 88 Сходимость NQS по энергии в зависимости от числа эпох обучения. По вертикали значение энергии на один спин, по горизонтали эпохи обучения. Пунктиром отмечено точное значение, полученное прямой диагонализацией гамильтониана. Слева общий вид, справа – сверхтонкая сходимость вблизи энергии основного состояния. Источник [CT17].

CNN над потенциалами

Другим интересным подходом является применение *Convolutional Neural Networks (CNN)* – специального класса глубоких нейронных сетей, разработанных для обработки изображений, к физическим потенциалам. Дело в том, что трехмерный энергетический потенциал можно представить как 2D изображение. Например, можно взять периодическую потенциальную энергию, созданную атомами в кристаллической решетке и рассмотреть, как будет вести себя электрон в таком потенциале.

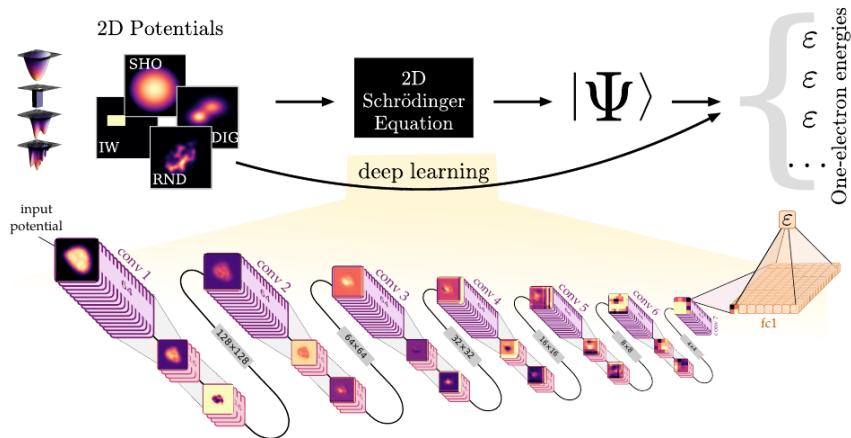


Fig. 89 Иллюстрация применения CNN на энергетических потенциалах. Источник [MST17].

Существует ограниченный набор частных случаев, когда эта задача может быть решена аналитически. Используя такие точные решения, мы составляем обучающую выборку, тренируем CNN, а потом можем применять ее к другим, более сложным потенциалам.

Решение уравнения Шредингера

Другим примером решения задачи об энергетических уровнях является применение глубокой нейронной сети, которая на вход принимает *Slater determinants*, *Jastrow factor* и другие известные способы приближенного описания квантовомеханических систем. А на выходе из такой глубокой нейронной сети получаем волновую функцию системы. Таким образом, получается объединить все преимущества хорошо изученных физических представлений с высокой экспрессивностью глубоких нейронных сетей [HSchatzleNoe20].

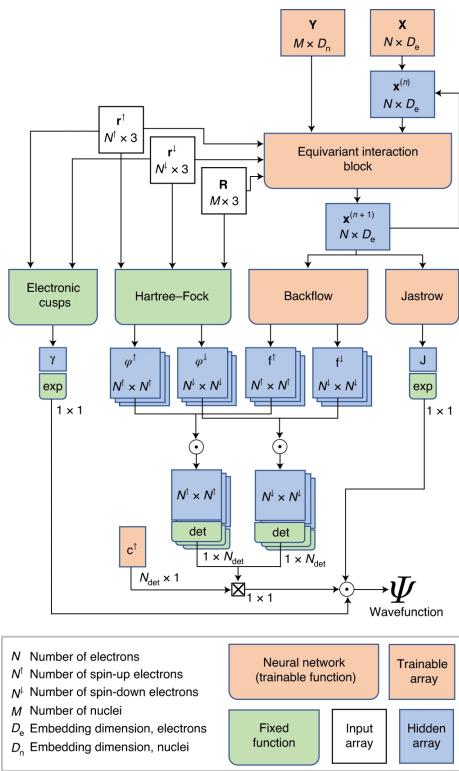


Fig. 90 Архитектура подхода *PauliNet*. Источник [[HSchatzleNoe20](#)].

Такой подход дает фантастическую точность по сравнению с другими подходами и выглядит очень перспективно, так как дополнительно позволяет получить очень хорошую масштабируемость из-за возможности снизить число *Slater determinants*.

RNN для моделирования волновых функций

Другой похожий на [NQS](#) подход – это использовать для аппроксимации волновой функции рекуррентные нейронные сети вместо машин Больцмана.



Fig. 91 Результаты RNN-модели для аппроксимации волновых функций. Можно получать не только значение энергии, но и ожидаемые значения других операторов, например, операторов спина или операторов спиновых корреляций. Источник [HAGH+20]. На графике (а) приведена зависимость энергии от эпохи обучения. На графике (б) приведены зависимости собственного значения операторов спиновых корреляций 40-го (центрального) спина с n -ым от номера n . На графике (с) приведено распределение значений оператора спина по длине цепочки спинов.

В данном подходе мы строим “последовательную” волновую функцию многочастичной квантовой системы, “подавая” туда по одной каждую частицу. А так как нейронная сеть – рекуррентная, мы по сути последовательно строим волновую функцию многочастичной системы.

Квантовая томография

Мы немного рассказывали о задаче квантовой томографии в [лекции про смешанные состояния](#). Правда, там привели пример использования метода максимизации правдоподобия для решения задачи восстановления состояния по измерениям. Оказывается, что для этих целей можно использовать глубокие нейронные сети. Один из таких подходов описан в [AMunozNK21]. Суть подхода в том, что используются генеративные сети с условием – *Conditional GAN* (C-GAN) для восстановления состояния. Для создания матрицы плотности используется генератор, а для вычисления ошибки относительно реальных результатов измерений – дискриминатор.

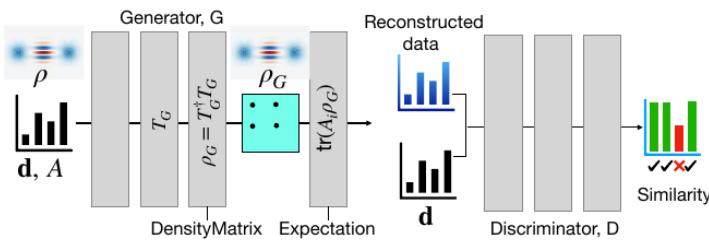


Fig. 92 Иллюстрация подхода C-GAN.

Заключение к подразделу

В целом, применение классических алгоритмов машинного обучения к задачам квантовой физики и квантовой химии – очень перспективная область. Мы еще немного коснемся этой темы, когда будем рассматривать решение задачи о собственных значениях гамильтонианов. Однако, эта тема выходит за рамки нашего курса.

Классические данные и квантовый алгоритм

Теперь перейдем к ситуации, когда мы рассматриваем классические данные (задачи классификации и регрессии) и применяем к ним машинное обучение, реализуемое на квантовом компьютере как набор операций над кубитами.

HHL

Один из самых известных алгоритмов квантового машинного обучения – это алгоритм **HHL** [[HHL09](#)], которому в нашем курсе [посвящена отдельная продвинутая лекция](#). Это алгоритм решения системы линейных уравнений за $\mathcal{O}(N \log N)$ операций. Так как к решению системы уравнений можно свести огромное число задач реального мира, этот алгоритм обладает огромным потенциалом. К сожалению, сегодня его практическое применение сильно ограничено несколькими вещами:

- необходимость эффективной генерации начального состояния – без этого достижения превосходства **HHL** невозможно;
- высокие требования к точности операций, а также необходимость в большом числе кубитов.

Этот алгоритм является примером чисто квантового алгоритма, где абсолютно все операции выполняются на квантовом компьютере.

Quantum k-NN

Другой пример полностью квантового алгоритма – это модификация классического алгоритма ближайшего соседа, с той лишь разницей, что для вычисления расстояний между точками в $\mathcal{V}(N)$ -мерном пространстве мы используем квантовый компьютер. Этот алгоритм называется *Quantum k-NN* (*Quantum k Nearest Neighbours*). Сегодня существует довольно много потенциально эффективных реализаций этого алгоритма, которые отличаются в основном тем, какая используется метрика расстояния в гильбертовом пространстве. Одна из реализаций [рассмотрена в продвинутой лекции нашего курса](#). К сожалению, практическое использование таких алгоритмов сегодня ограничено тем, что для них требуется эффективная квантовая память – *quantum Random Access Memory*.

Заключение к подразделу

Сегодня мы живем в так называемой *NISQ* (*Noise Intermedium Scale Quantum*) эпохе, то есть когда у нас есть квантовые компьютеры лишь ограниченного размера и с относительно большим уровнем шумов. А чисто квантовые алгоритмы квантового машинного обучения очень требовательны именно к точности вычислений. Та же в таких алгоритмах большой проблемой является постоянная необходимость сложной операции перевода классических данных в квантовые. В теории, эта проблема исчезнет с появлением *qRAM*, однако сегодня такая “квантовая память” еще не существует и даже пока нет однозначного понимания того, как именно ее делать. Все это сильно ограничивает потенциал полностью квантовых подходов. Далее мы рассмотрим более перспективный в *NISQ*-эпоху подход – гибридное квантово-классическое машинное обучение.

Гибридное обучение

Большая часть из оставшихся лекций нашего курса будет посвящена как раз гибридным методам квантового машинного обучения. В этих методах часть алгоритма реализуется в виде квантовых вентилей, а часть выполняется на классическом компьютере.

VQC

Variational Quantum Circuits, или просто вариационные схемы, – это одно из центральных понятий в гибридном квантово-классическом обучении. Основная идея заключается в том, что мы используем квантовую операцию, которая задается каким-то классическим параметром. Обычно это одна или несколько операций “вращений” на сфере Блоха, про которые [мы говорили в лекции про квантовые гейты](#). При этом вариация классического параметра осуществляется на классическом компьютере, например, при помощи градиентного спуска.



Fig. 93 Схема работы вариационной квантовой схемы.

В нашем курсе теме вариационных схем будет посвящено несколько лекций:

- вариационные квантовые схемы
- градиенты **VQC**
- продвинутая лекция по градиентам **VQC**

Также мы часто будем возвращаться понятию VQC в теме квантовых нейросетей, а также в блоке про решение задачи собственных значений при помощи алгоритма *Variational Quantum Eigensolver*.

Квантовые нейросети

Квантовые “нейронные сети” – это как раз пример, когда мы комбинируем вариационные слои вместе с обычными слоями нейронных сетей.

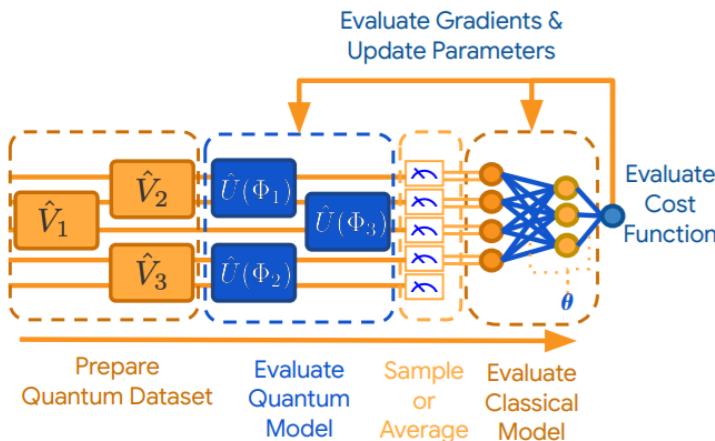


Fig. 94 Пример гибридной квантовой нейросети. Источник [BVM+21].

На картинке выше приведен пример комбинации вариационных квантовых схем и классических слоев обычных глубоких нейронных сетей. При этом оптимизация параметров выполняется при помощи единого процесса *backpropagation*, с той лишь разницей, что для классических и квантовых слоев немного по-разному вычисляется градиент для параметров конкретного слоя. Теме квантовых нейронных сетей в нашем курсе будет посвящен целый блок, где будет, в том числе, рассказано еще и об аналоге сверточного слоя с использованием **VQC**.

Квантовые ядра

Квантовые ядра и квантовый алгоритм *Support Vector Machine (SVM)* – это другой пример, как можно объединить квантовые гейты с классическими алгоритмами. Математике “под капотом” классического SVM у нас будет посвящена отдельная лекция, но если кратко, то суть там в том, что решение задачи об оптимальной разделяющей гиперплоскости (другими словами, об оптимальной классификации) можно выразить через

скалярные произведения точек обучающей выборки. Причем не обязательно в исходном пространстве, а в любом гильбертовом пространстве. А как мы помним, квантовые гейты как раз представляют собой операции в гильбертовом пространстве, причем экспоненциально большого размера. Так и появляется идея гибридного SVM – мы переписываем скалярные произведения точек как результат измерения параметризованных квантовых схем, где параметры это и есть компоненты каждой из точек. А дальше уже применяем классические методы решения оптимизационной задачи. Этой теме у нас будет посвящена отдельная лекция.

Оптимизация энергии

Еще один интересный и перспективный класс задач, которые можно решать гибридными методами – это оптимизация энергии гамильтониана. Может показаться, что это еще что-то из области квантовой физики, но на самом деле в блоке о проблемах квантового машинного обучения мы покажем, как к задаче поиска основного состояния системы, описанной квантовым оператором – гамильтонианом, можно свести огромное число задач реального мира. Например, это задача Коммивояжера, очень важная в области логистики, или задача о нахождении энергий электронов, которая играет важную роль в разработке лекарств или создания новых материалов.

Заключение

В этой лекции мы познакомились с тем, каким может быть квантовое машинное обучение. Мы узнали, что бывают:

- классические алгоритмы над квантовыми данными;
- квантовые алгоритмы над классическими данными;
- гибридные алгоритмы.

Именно гибридные алгоритмы наиболее перспективны в NISQ-эпоху, и именно им будет посвящена большая часть оставшихся лекций курса.

Алгоритм HHL

Сегодня пришла пора поговорить о знаменитом алгоритме Харроу, Хассидима и Ллойда, более известном как HHL-алгоритме, способном решать системы линейных уравнений.

Очень надеюсь, что к данному занятию у вас уже есть представление об алгоритме фазовой оценки (QPE), использующем обратное квантовое преобразование Фурье, на котором и базируется HHL. Глубокое понимание всех тонкостей этого алгоритма потребует от вас уверенного владения математическим аппаратом. За детальным описанием вы всегда можете обратиться к статьям [[DHM+18](#)], [[HHL09](#)], и [[HZL+17](#)]. Приготовьтесь потратить время и умственные ресурсы, если алгоритм вас зацепит и вы решите в нем как следует покопаться. Мы же поможем вам заинтересоваться, рассмотрим основные принципы и небольшой пример.

Note

Именно HHL-алгоритм произвел настоящую революцию в области квантового машинного обучения. Ведь решение систем линейных уравнений так или иначе находится “под капотом” почти любого известного алгоритма машинного обучения. И действительно:

- классические линейная и логистическая регрессия сводятся именно к этой задаче;
- задача SVM может быть переформулирована в терминах решений систем линейных уравнений;
- задача нахождения обратной матрицы (часто используется в глубоком обучении) внутри обычно решается через решение линейной системы;

И это только малая часть примеров!

Так что знакомство с QML не будет полным без ознакомления с этим прекрасным, но очень сложным алгоритмом!



Fig. 95 Сет Ллойд, профессор МИТ и один из создателей HHL-алгоритма

Задача

Представим обычную систему линейных уравнений:

$$\begin{aligned} \left[\begin{array}{ll} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{array} \right] \end{aligned}$$

Что в операторной форме можно переписать как:

$$[\mathbf{A}\vec{x}] = \vec{b}$$

где \mathbf{A} – эрмитова матрица.

Мы будем решать задачу на квантовом компьютере, то нам нужно перейти к квантовым состояниям:

$$[\mathbf{A}|x\rangle = |b\rangle]$$

Чтобы найти искомый вектор $|x\rangle$, все что нам по сути нужно сделать – это найти обратный к \mathbf{A} оператор (обозначаемый \mathbf{A}^{-1}), который находится из равенства:

$$[\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = I]$$

Распишем применительно к нашей задаче поиска вектора $|b\rangle$:

$$\begin{aligned} [\mathbf{A}|x\rangle = |b\rangle \quad \mathbf{A}^{-1}\mathbf{A}|x\rangle = \mathbf{A}^{-1}|b\rangle \quad |x\rangle = \mathbf{A}^{-1}|b\rangle] \end{aligned}$$

Оказывается, что и это все можно провернуть с помощью известных квантовых преобразований.

Принципиальный вид нашей схемы представлен следующим образом:



Fig. 96 Квантовая схема, реализующая алгоритм HHL

В нижний регистр загружается вектор $|b\rangle$, средний и нижний регистры участвуют в фазовой оценке, а верхний дополнительный кубит нужен для так называемого вращения, обусловленного собственными значениями. Давайте разбираться.

Реализация HHL

Для начала мы должны подготовить наши регистры по всем квантовым законам: $|b\rangle$ и $|x\rangle$ должны быть пронормированы, а оператор \mathbf{A} должен быть эрмитовым. Надеемся, что Вы помните про ортонормированный базис, сферу Блоха, комплексное представление векторов $|0\rangle$ и $|1\rangle$... Если нет, то обратитесь к предыдущим разделам курса.

Мы будем использовать оператор $\langle U = e^{iAt} \rangle$, и нужно, чтобы он был обратим — для этого $\langle A \rangle$ должна быть эрмитовой.

Note

В случае, когда $\langle A \rangle$ не является эрмитовой, нужно перейти к эрмитовой матрице $\langle C \rangle$:

$$\begin{aligned} & \langle C \rangle = \begin{pmatrix} 0 & A^\dagger \\ A & 0 \end{pmatrix} \end{aligned}$$

И рассматривается задача $\langle C | \vec{v} \rangle = \left(\begin{array}{c} \langle b \rangle \\ 0 \end{array} \right)$ для того, чтобы найти решение $\langle y \rangle = \left(\begin{array}{c} \langle 0 \rangle \\ \langle x \rangle \end{array} \right)$

Вспомним, что эрмитову матрицу $\langle A \rangle$ можно представить в виде суммы собственных векторов, умноженных на собственные значения, т.е. в виде спектрального разложения:

$$\begin{aligned} & \langle A \rangle = \sum_{j=0}^{N-1} \lambda_j \langle u_j \rangle \langle u_j | \end{aligned}$$
$$\begin{aligned} & \langle A^{-1} \rangle = \sum_{j=0}^{N-1} \lambda_j^{-1} \langle u_j \rangle \langle u_j | \end{aligned}$$

Тогда вектор $\langle b \rangle$ можно представить через собственные векторы $\langle A \rangle$:

$$\langle b \rangle = \sum_{j=0}^{N-1} b_j \langle u_j \rangle$$

Чтобы понять, почему это тоже ключевой момент, давайте вспомним, что значит собственный вектор и собственное значение матрицы.

Note

Собственным вектором $\langle u \rangle$ оператора $\langle A \rangle$ называется такой ненулевой вектор, для которого выполняется:

$$\langle A | u \rangle = \lambda u$$

λ — собственное значение оператора $\langle A \rangle$.

Таким образом, искомый вектор $\langle x \rangle$ — не что иное, как:

$$\langle x \rangle = A^{-1} \langle b \rangle = \sum_{j=0}^{N-1} \lambda_j^{-1} b_j \langle u_j \rangle$$

Итак, фазовая оценка. Мы применяем к кубитам второго регистра матрицы Адамара, тем самым приводим их в суперпозицию. Следом запускаем оператор $\langle U \rangle$:

$$\langle U | = e^{iAt} = \sum_{j=0}^{N-1} e^{i\lambda_j t} \langle u_j \rangle$$

Для того, чтобы узнать собственное значение оператора $\langle U \rangle$, получения фазы (Quantum Phase Estimation – QPE), результатом которого получится следующее состояние:

$$\langle QPE(U) | 0 \rangle = \frac{1}{\sqrt{2^m}} \langle 0 | \langle 1 | \dots \langle 1 | \langle 0 | + e^{i2\pi t} \langle 0 | \langle 1 | \dots \langle 1 | \langle 1 |$$
$$\langle 0 | \dots \langle 0 | + e^{i2\pi t} \langle 0 | \langle 1 | \dots \langle 1 | \langle 0 | + \dots + e^{i2\pi t} \langle 0 | \langle 1 | \dots \langle 1 | \langle 1 |$$

Параметр t это нормировочная константа в случае $\langle U = e^{iAt} \rangle$:

$$e^{i2\pi t} = \frac{\lambda}{\sqrt{\lambda^2 + 1}}$$

Параметр t подбирается с учетом того, что на выходе алгоритма QPE собственные значения λ нормализуются к виду $0 \leq \lambda \leq 1$ и обычно мы располагаем ограниченным числом кубитов, которое можно использовать для аппроксимации.

Алгоритм обратного квантового Фурье переводит фазу в конкретный вектор.

Принципиальная схема QPE выглядит следующим образом:

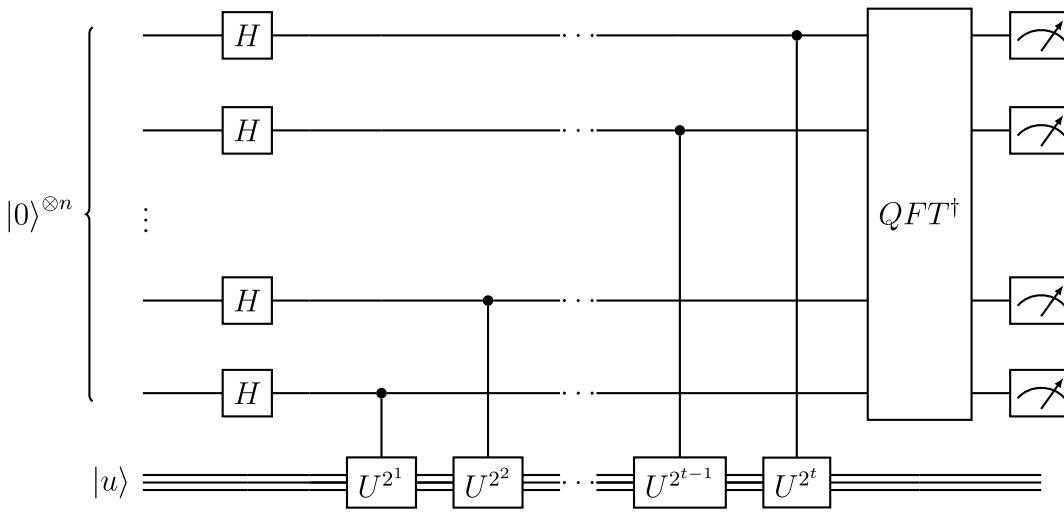


Fig. 97 Схема алгоритма QPE

Итак, мы подготовились, вспомнили много хорошего, теперь пошагово распишем наш алгоритм.

Стартуем мы со следующим состоянием:

$|0\rangle^{\otimes n}$

Т.е. наше состояние будет храниться в трех регистрах, в каждом из которых содержится столько кубитов, сколько нужно для решения задачи.

1. Применение QPE с использованием преобразования e^{iAt} , после чего мы получим собственное

значение оператора $\langle A \rangle$ во втором регистре:
 $\sum_{j=0}^{N-1} b_j |\lambda_j\rangle$

2. Поворачиваем первый кубит (с индексом $\langle A \rangle$) с помощью специальный оператор вращения R :

$\{|\lambda_j\rangle\} |1\rangle$

где $\langle C \rangle$ – константа, которая должна быть меньше минимального из лямбда: $|(C| < |\lambda_{\min}|)$
 [Почему?].

Переворачиваем первый кубит $\langle 0\rangle$:
 $\sum_{j=0}^{N-1} b_j |\lambda_j\rangle |1\rangle$

$\{|\lambda_j\rangle\} |1\rangle$

3. Применяем $(QPE)^{\dagger}$ (т.е. обратное получение фазы) и получаем следующее состояние:

$\{|\lambda_j\rangle\} |1\rangle$

$\{|\lambda_j\rangle\} |0\rangle$

В конце мы измеряем верхний кубит и если получаем единицу, то знаем, что в нижнем регистре хранится

искомый $\langle A \rangle$ с учетом нормировки:

Пример

Рассмотрим небольшой, но удобный пример. Удобный в том отношении, что, вообще говоря, алгоритм HHL имеет определенное приближение. Если собственные значения не представимы в бинарной форме, то о 100% точности говорить не приходится. Мы также опустим ряд вопросов, связанных с подбором параметров и количества кубитов второго регистра. Главное сейчас – понять, что происходит, и для этого наша матрица $\langle A \rangle$ эрмитова, все условия подобраны, а преобразования точны. Стоит помнить, что знать заранее значение собственных векторов и собственных значений нам совершенно не обязательно – это нужно лишь для наглядности.

Итак, пусть задача выглядит так :

```
\begin{aligned}
& \& A = \left( \begin{array}{cc} 1 & \frac{3}{5} \\ 0 & 1 \end{array} \right) \\
& \& \langle A \rangle = \left( \begin{array}{cc} 1 & \frac{3}{5} \\ \frac{3}{5} & 1 \end{array} \right)
\end{aligned}
```

Собственные значения и соответствующие собственные векторы:

```
\begin{aligned}
& \& \lambda_0 = \frac{2}{5}\pi, |\lambda_0\rangle = \left( \begin{array}{c} -1 \\ 1 \end{array} \right) \\
& \& \lambda_1 = \frac{8}{5}\pi, |\lambda_1\rangle = \left( \begin{array}{c} 1 \\ 1 \end{array} \right)
\end{aligned}
```

Зададим параметр t и проанализируем фазу:

```
\begin{aligned}
& \& \psi = 2\pi \frac{\sqrt{5}\sqrt{6}}{10} \lambda_0 t \\
& \& \text{если } \lambda_0 t / (2\pi) \leq \frac{1}{8} \text{, то } \psi = \frac{\lambda_0 t}{2\pi} \pi
\end{aligned}
```

Как мы видим, для перевода угла $\langle \psi \rangle$ в векторную форму, нам понадобятся три кубита. После преобразования QPE мы имеем следующее состояние:

$$\begin{aligned} & \left\langle \begin{aligned} & \text{\&\{QPE(|}\rangle_a|\rangle_r|b\rangle_m) = \sum_{j=0}^1 \frac{1}{\sqrt{2}} \left(|0\rangle_a |001\rangle_r |u_0\rangle_m + \right. \right. \\ & \left. \left. |0\rangle_a |100\rangle_r |u_1\rangle_m \right) \end{aligned} \right\rangle \end{aligned}$$

Подберем константу $\langle C \rangle$ (как мы помним, она должна быть меньше наименьшего из собственных чисел) и произведем вращение:

$$\begin{aligned} & \left\langle \begin{aligned} & \text{\&\{split\} \begin{aligned} & \text{\&\{aligned\} \&\frac{1}{\sqrt{2}} \left(|0\rangle_a |001\rangle_r |u_0\rangle_m + \right. \right. \\ & \left. \left. |0\rangle_a |100\rangle_r |u_1\rangle_m \right) \end{aligned} \end{aligned} \right\rangle \end{aligned}$$

В конце мы производим измерение верхнего кубита (с индексом $|a\rangle$) и при получении единицы можем быть уверены, что нижний регистр содержит искомое решение с учетом нормировки:

$$\begin{aligned} & \left\langle \begin{aligned} & \text{\&\{split\} \begin{aligned} & \text{\&\{aligned\} \frac{1}{\sqrt{2}} \left(|0\rangle_a |001\rangle_r |u_0\rangle_m + \right. \right. \\ & \left. \left. |0\rangle_a |100\rangle_r |u_1\rangle_m \right) \end{aligned} \end{aligned} \right\rangle \end{aligned}$$

Quantum K-nearest neighbor

Введение

Если вы занимались машинным обучением, то, скорее всего, знакомы с классическим алгоритмом $\langle k \rangle$ ближайших соседей. Он относительно прост, применяется как в задачах классификации, так и в регрессии. Кстати, классический knn можно вспомнить обратившись к [лекции](#) от ODS по классическому машинному обучению.

Давайте немножко вспомним задачу классификации с использованием классического $\langle kNN \rangle$ алгоритма:

У нас есть $\langle \text{vec}(x) \in \{0, 1\}^N \rangle$ – **тестовый образец**, а также **тренировочные образцы** – это набор векторов $\langle \text{vec}(v_i) \in \{0, 1\}^N \rangle$, в котором каждый вектор уже размечен. И наша задача подобрать правильную метку тестовому образцу.

Тогда мы пройдём следующие шаги:

1. Вычислим похожесть между тестовым образцом и **каждым** тренировочным образцом.
2. Найдем $\langle k \rangle$ ближайших к тестовому образцу соседей.
3. Подсчитаем количество представителей для каждого класса и приписываем метку самого часто встречающегося класса к тестовому образцу.

Самой трудозатратным шагом является вычисление расстояния от тестового образца к каждому тренировочному. Также и в квантовой версии алгоритма.

Note

На текущий момент разработано несколько разных версий квантового алгоритма поиска ближайших соседей. Есть версия основанная расстоянии Хэмминга [\[LLKG21\]](#):

Расстояние Хэмминга между векторами $\langle \text{vec}(x) \rangle$ и $\langle \text{vec}(v_i) \rangle$:

$$d_i = |\text{vec}(x) - \text{vec}(v_i)| = \sum_{j=1}^N (x_j \oplus v_{ij})$$

Но в данной работе мы обратим внимание на версию, которая вычисляет fidelity между двумя векторами состояниями.

Пусть задано Гильбертово пространство $\langle n \rangle$ кубитов, размерности $\langle N = 2^n \rangle$. Вектор $\langle |\psi\rangle \in H \rangle$ – это тестовое состояние, для которого нам нужно определить метку.

Пусть $\langle \{\langle \phi_i \rangle | i \in \{0, \dots, M-1\} \} \subset H \rangle$ – это набор тренировочных состояний, для которых мы знаем их метки. $\langle M = 2^m, m \in \mathbb{Z} \rangle$

Определим fidelity между тестовым состоянием и $\langle j \rangle$ -тым тренировочным $\langle |\phi_j\rangle$ как

$$F_j = F(\psi, \phi_j) = |\langle \psi | \phi_j \rangle|^2$$

В свою очередь $\langle F = [F_0, F_1, \dots, F_{M-1}] \rangle$ – это таблица fidelity значений между тестовым состоянием и **каждым** из тренировочных.

Заметим, что задача нахождения $\langle k \rangle$ ближайших соседей сводится к задаче нахождения $\langle k \rangle$ максимумов значений fidelity из таблицы $\langle F \rangle$. Для этого мы должны реализовать оракула

$\langle O_{\{y, A\}} | j \rangle = \langle j | f_{\{y, A\}}(j) \rangle$
 где $f_{\{y, A\}}$ - это булева функция определённая как
 $\begin{cases} 1 : F_j > F_y \text{ и } j \notin A, \\ 0 : \text{otherwise} \end{cases}$

Алгоритм

Далее мы представим алгоритм, представленный в работе [BAG21].

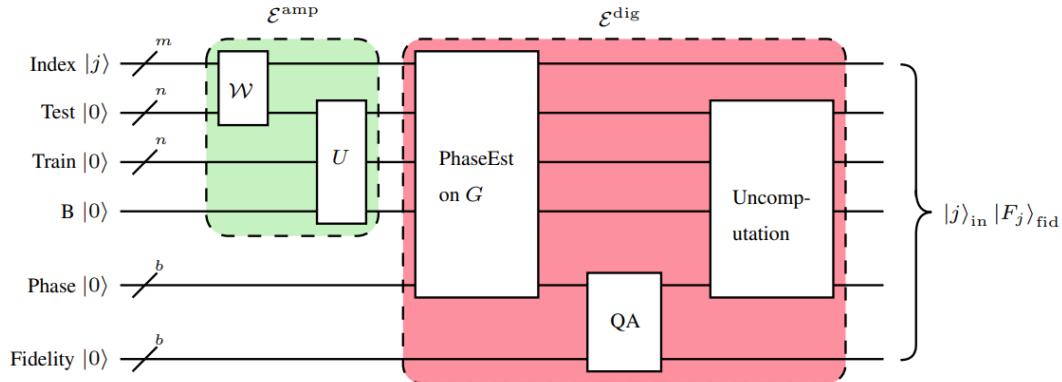


Fig. 98 Принципиальная схема QkNN алгоритма. Взято из работы [BAG21].

Квантовый алгоритм поиска $|k\rangle$ ближайших соседей будет состоять из двух основных шагов:

- Используй оракул $\langle O_{\{y, A\}} |$ (для алгоритма Гровера) мы находим $|k\rangle$ состояний $\langle \psi_{|k\rangle} |$, для которых значение fidelity с тестовым состоянием максимально.
- Найти преобладающую метку среди $|k\rangle$ найденных состояний и присвоить её тестовому состоянию.

Самой нетривиальной задачей для нас будет построение оракула $\langle O_{\{y, A\}} |$.

- Вначале нужно составить оператор $\langle \mathcal{F}_j |$, который выполняет преобразование вида:

$$\langle \mathcal{F}_j | = \sum_{i=0}^{M-1} \langle i | \psi_i \rangle \langle \psi_i | \mathcal{F}_j |$$
 для $j \in \{0, \dots, M-1\}$. $\langle \mathcal{F}_j |$ - это одно из базисных состояний вычислительного базиса (выражающее двоичное представление числа $|F_j\rangle$).
 - Выполняется преобразование: $\langle x | \mathcal{F}_j | = \langle j | \psi_i \rangle \langle \psi_i | \mathcal{F}_j |$. В амплитуду состояния $\langle \psi_i |$ закодирована информация о числе $|F_j\rangle$. Делается это с помощью Swap test.
 - Swap test это применение контролируемой операции $\langle \text{Swap} |$, которым можно пользоваться для того, чтобы статистически определять fidelity: $|\langle \mathcal{F}_j | \psi_i \rangle|^2 = |\langle \psi_i | \mathcal{F}_j ||^2$ между двумя производными чистыми состояниями $\langle \psi_i |$ и $\langle \mathcal{F}_j |$.

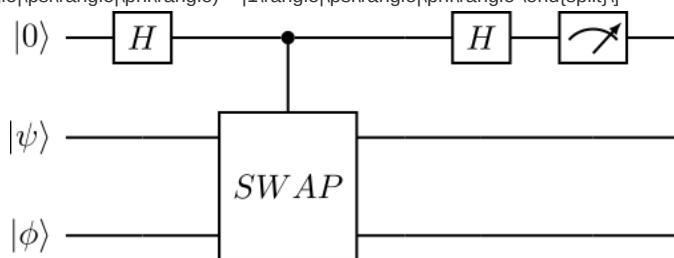


Fig. 99 Схема Swap test

- Выполняется преобразование $\langle x | \mathcal{F}_j | = \langle j | \psi_i \rangle \langle \psi_i | \mathcal{F}_j |$
 И тогда $\langle \mathcal{F}_j | = \langle x | \mathcal{F}_j |$.
- Берём 2 пары регистров $|i_1, f_1; i_2, f_2\rangle$. Инициализируются они в форме $|i_1\rangle |f_1\rangle |i_2\rangle |f_2\rangle$
 Применяется $\langle \mathcal{F}_j |$ на каждой паре:

$$\langle \mathcal{F}_j | = \langle i_1 | \mathcal{F}_j | + \langle i_2 | \mathcal{F}_j |$$

$$= \langle i_1 | F_j | i_1 \rangle | f_1 \rangle + \langle i_2 | F_j | i_2 \rangle | f_2 \rangle$$
- Теперь информация закодирована в регистры и нам нужно реализовать функцию $\langle f_{\{y, A\}} |$. И пусть $\langle \mathcal{C} |$ это оператор, реализующий $\langle f_{\{y, A\}} |$:

$$\langle \mathcal{C} | = \langle i_1 | f_{\{y, A\}} | i_2 \rangle | f_2 \rangle$$

$$= \langle i_1 | F_j | i_1 \rangle | f_1 \rangle | f_2 \rangle$$

Теперь займёмся вопросом конструирования оракула $\langle O_{\{y, A\}} |$. Просьба держаться за ваши кресла.

Вначале мы подготовим состояния. Но чтобы это сделать нам нужны оракулы $\langle \mathcal{V}, \mathcal{W} \rangle$. Как их имплементировать указано в статье, которая указывалась выше.

$\langle \mathcal{V}_j, \mathcal{W}_j \rangle = |\psi_j\rangle\langle \phi_j|$

для всех $j \in \{0, \dots, M-1\}$.

1. Инициализируем 4 регистра $(\mathbf{l}), (\mathbf{tr}), (\mathbf{tst}), (\mathbf{B})$ с соответствующим количеством кубитов в каждом $(m), (n), (m), (n)$, где $n = \text{Vec}(\mathbf{N}), m = \text{Vec}(\mathbf{M})$

$$\langle \mathcal{V}_j, \mathcal{W}_j \rangle = |\psi_j\rangle\langle \phi_j| = |\phi_j\rangle\langle \phi_j|_B$$
2. Применяем $\langle \mathcal{W}_j, \mathcal{W}_j \rangle$

$$\langle \mathcal{W}_j, \mathcal{W}_j \rangle = |\otimes_{i=1}^m \langle \phi_j | \rangle_i \otimes_{i=1}^n \langle \phi_j | \rangle_i = |\phi_j\rangle\langle \phi_j|_B$$
3. Применяем $\langle \mathcal{V}_j, \mathcal{W}_j \rangle$

$$\langle \mathcal{V}_j, \mathcal{W}_j \rangle = |\psi_j\rangle\langle \phi_j| = |\psi_j\rangle\langle \phi_j|_B$$
4. Применяем swap test между тренировочным регистром (\mathbf{tr}) и тестовым регистром (\mathbf{tst}) , а регистр (\mathbf{B}) будет выступать в качестве контрольного

$$|\frac{1}{\sqrt{2}}(|\psi_j\rangle\langle \phi_j| + |\phi_j\rangle\langle \psi_j|)|\rangle_B = |\frac{1}{\sqrt{2}}(|\psi_j\rangle\langle \phi_j| + |\phi_j\rangle\langle \psi_j|)|\rangle_B$$
- Определим (U) как унитарное преобразование, которое объединяет шаги 3-4. Кстати, если мы сейчас произведём измерение регистра (\mathbf{B}) , то будем иметь $|\frac{1}{\sqrt{2}}(1 - F_j)|^2$

$$|\text{begin}\{\text{split}\} |\mathcal{P}(\mathbf{B}) \text{ на } |\frac{1}{\sqrt{2}}(1 - F_j)|^2 \text{ будем иметь} |\text{end}\{\text{split}\}|$$
- На этом шаге информация о (fidelity) теперь закодирована в амплитуды. Теперь же мы должны перевести (fidelity) из амплитуды в число.
5. Теперь мы будем конструировать новый гейт G . Вообще говоря, он описан в работе [MKF19], где вы можете подробнее с ним познакомиться

$$|\mathcal{G} = |\mathcal{O}_{\text{tr}}, \mathcal{S}_{\text{tr}}, \mathcal{B}_{\text{tr}}\rangle\langle \mathcal{W}_j |, |\mathcal{I}_{\text{tr}}\rangle\langle \mathcal{S}_{\text{tr}}|, |\mathcal{W}_j\rangle\langle \mathcal{U}_{\text{tr}, \text{tst}, \mathbf{B}}|, |\mathcal{Z}_{\text{B}}\rangle\langle \mathcal{Z}_{\text{B}}|$$
 где (\mathcal{Z}_{B}) – это действие гейта (Z) на регистре (\mathbf{B}) , $(\mathcal{S}_{\text{tr}} = I - 2|\psi_j\rangle\langle \psi_j|)$.
6. Текущее состояние $(|\psi_j\rangle\langle \phi_j|_{\text{tr}, \text{tst}, \mathbf{B}})$ может быть представлено в виде композиции двух состояний $(|\psi_j\rangle\langle \phi_j|_{\text{tr}, \text{tst}, \mathbf{B}} = |\psi_j\rangle\langle \phi_j|_{\text{tr}, \text{tst}, \mathbf{B}} \otimes |\phi_j\rangle\langle \phi_j|_{\text{tst}, \mathbf{B}})$
7. Теперь применяем алгоритм QPE (Quantum Phase Estimation), чтобы перевести значение фазы $(\langle \theta_j \rangle)$ в числовое представление

$$|\text{QPE}(\psi_j, \phi_j) = |\frac{1}{\sqrt{2}}(\langle \theta_j \rangle + i\langle \theta_j \rangle)|\rangle = |\frac{1}{\sqrt{2}}(\langle \theta_j \rangle + i\langle \theta_j \rangle)|\rangle$$
8. Применяем алгоритм квантовой арифметики

$$|\text{QuantumArith}(\phi_j) = |\psi_j, \text{AE}\rangle\langle \phi_j|$$
9. Обнуляем регистры $(\mathbf{ph}, \mathbf{tr}, \mathbf{tst}, \mathbf{B})$ и получаем $(|\psi_j\rangle\langle \phi_j|_{\text{fid}})$ На самом деле шаги 5-9 составляют оператор $(\text{xi}[\text{dig}])$, который мы упоминали ранее.
10. Теперь применяем оператор $(\langle \mathcal{F} \rangle)$

$$|\langle \mathcal{F} | = |\mathcal{F}_1\rangle\langle \mathcal{F}_1| + |\mathcal{F}_2\rangle\langle \mathcal{F}_2|$$
11. Добавим кубит (Q_1) для выполнения сравнения

$$|\text{begin}\{\text{split}\} |\mathcal{P}(Q_1) \text{ для выполнения сравнения} |\text{begin}\{\text{cases}\} |\mathcal{A}\rangle\langle \mathcal{B}| |\text{begin}\{\text{cases}\} |\mathcal{A}\rangle\langle \mathcal{B}| : a > b, ||$$

$$|\text{begin}\{\text{else}\} |\mathcal{P}(Q_1) \text{ для выполнения сравнения} |\text{begin}\{\text{cases}\} |\mathcal{A}\rangle\langle \mathcal{B}| |\text{begin}\{\text{cases}\} |\mathcal{A}\rangle\langle \mathcal{B}| : a < b, ||$$

$$|\text{begin}\{\text{split}\} g(j) = |\text{begin}\{\text{cases}\} 1 : F_j > F_y, || 0 : F_j \leq F_y, |\text{end}\{\text{cases}\} |\text{end}\{\text{split}\}|$$
 По кубиту (Q_1) мы сможем распознать все индексы (j) для которых $(F_j > F_y)$.
12. Обнуляем регистры $(\mathbf{f1}), (\mathbf{f2})$.
13. Добавляем ещё один кубит (Q_2) для каждого $(j \in \text{Vec}(\mathbf{A}))$, применяя гейт $(\text{D}(\langle \mathcal{F}_j \rangle))$

$$|\text{begin}\{\text{split}\} |\mathcal{D}(\langle \mathcal{F}_j \rangle) \text{ для каждого } j \in \text{Vec}(\mathbf{A}) : j \neq i_l, ||$$

$$|\text{begin}\{\text{cases}\} |\mathcal{D}(\langle \mathcal{F}_j \rangle) \text{ для каждого } j \in \text{Vec}(\mathbf{A}) : j = i_l, ||$$
 На индексах регистра (\mathbf{f}_2) в результате получим состояние $(|\mathcal{F}_j\rangle_{\mathbf{f}_2})$
14. О да... Теперь мы добавляем ещё один кубит (Q_3) . Применяем гейт (X) на кубите (Q_2) и гейт Тoffoli с контролирующими (Q_1, Q_2) и целевой (Q_3)

$$|\text{begin}\{\text{split}\} |\mathcal{T}(\langle \mathcal{F}_1 \rangle, \langle \mathcal{F}_2 \rangle) \text{ и целевой } (\langle \mathcal{F}_3 \rangle) |\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4\rangle\langle \mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4|$$
15. Обнуляем все регистры, кроме (Q_3)

$$|\mathcal{O}_{\text{y}, \mathbf{A}}\rangle\langle \mathcal{O}_{\text{y}, \mathbf{A}}|$$
 Что ж, вот мы и построили преобразование $(\mathcal{O}_{\text{y}, \mathbf{A}})$ которое так хотели

О блоке “Вариационные схемы”

Этот блок включает в себя только вариационные квантовые схемы (VQC).

Вариационные квантовые схемы

Описание лекции

На этой лекции мы впервые познакомимся непосредственно с квантовым машинным обучением. Лекция расскажет:

- в чем заключается идея квантово-классического обучения;

- что такое вариационное машинное обучение;
- как устроена вариационная квантовая схема и как закодировать в нее данные.

Введение

В течение всего курса мы будем говорить преимущественно о комбинированном квантово-классическом машинном обучении, построенном на базе вариационных квантовых схем. Именно он является наиболее перспективным в эпоху [NISQ](#). Давайте забежим немного вперед и посмотрим, как выглядит типичный цикл такого обучения.

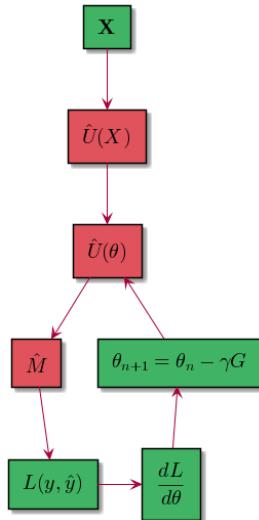


Fig. 100 Пример цикла смешанного обучения

Красным цветом на диаграмме помечены блоки, выполняемые на квантовом устройстве, зеленым – те, что считает обычный компьютер. Квантовая часть, которая включает в себя операторы $\hat{U}(X)$ и $\hat{U}(\theta)$, а также измерение наблюдаемой \hat{M} , называется вариационной квантовой схемой. Именно ей посвящена данная лекция.

Но сначала сделаем шаг назад и обсудим в общих чертах идеи, которые лежат в основе квантово-классического обучения.

Квантово-классическое обучение

Основная идея квантово-классического обучения заключается в том, что в эпоху Noise Intermediate Scale Quantum (*NISQ*) у нас нет больших работающих квантовых компьютеров и квантовой памяти. Это сильно ограничивает применение алгоритмов, более быстрых, чем классические. Нам остается только “встраивать” квантовые схемы в классический цикл обучения.

В этом случае мы разделяем классическую и квантовую части алгоритма. Выполняем предварительную обработку и подготовку данных на классическом компьютере, затем “прогоняем” их через квантовую схему. Она должна возвращать нам “классические” данные, а значит, должна включать в себя набор последовательно применяемых операторов и измерение. Со стороны классического компьютера такая схема выглядит как “черный ящик” или “оракул”, с которым уже можно работать. Например, можно варьировать параметры схемы таким образом, чтобы она “обучалась” так же, как “обучаются” нейронные сети или другие алгоритмы классического машинного обучения.

Давайте более детально посмотрим на квантовую схему, которую можно использовать в таком подходе.

Вариационные квантовые схемы

В основе вариационных квантовых схем – **Variational Quantum Circuits (VQC)** – лежит простая идея. Чтобы проиллюстрировать ее, давайте сделаем схему, которая включает в себя набор унитарных операторов и переводит классические данные \mathbf{X} и параметры θ в квантовое состояние $|\Psi\rangle$. Затем будем варьировать наше состояние $|\Psi\rangle$, меняя параметры θ так, чтобы при его измерении в каком-либо базисе мы получали нужный результат, например, результат классификации входных данных \mathbf{X} .

Разберем написанное выше на примере.

Аппроксимация матрицы Паули **X** оператором поворота

Из предыдущих лекций мы помним, что оператор \hat{X} работает как квантовый аналог оператора NOT – инвертирует значение кубита.

```
import pennylane as qml

dev = qml.device("default.qubit", 1)

@qml.qnode(dev)
def simple_x_gate():
    qml.PauliX(0)
    return qml.expval(qml.PauliZ(0))

print(f"simple_x_gate() = {simple_x_gate()}")
print(simple_x_gate.draw())
```

```
simple_x_gate() = tensor(-1., requires_grad=True)
0: ─X─ (Z)
```

С другой стороны, у нас есть оператор $\hat{RX}(\phi)$, который “вращает” состояние нашего кубита вокруг оси X на угол ϕ . Сделаем параметризованную схему с одним параметром:

```
@qml.qnode(dev)
def vqc(phi):
    qml.RX(phi, wires=[0])
    return qml.expval(qml.PauliZ(0))
```

Теперь попробуем подобрать ϕ так, чтобы параметризованная схема работала точно так же, как оператор \hat{X} :

```
for i in [1.0, 2.0, 3.0, 3.14159265359]:
    print(f"Try {i}: {vqc(i)}")
```

```
Try 1.0: 0.5403023058681398
Try 2.0: -0.4161468365471423
Try 3.0: -0.9899924966004454
Try 3.14159265359: -1.0
```

Получается, что нужный нам угол ϕ составляет ровно π . Это логично, мы могли бы это легко понять из простейших соображений линейной алгебры, либо взглянув на сферу Блоха. Но целью этого примера было проиллюстрировать работу VQC.

Наш процесс “подбора” параметра ϕ крайне примитивный, на следующих лекциях мы рассмотрим, как посчитать градиент параметров квантовой схемы и задействовать всю мощь известных на сегодня методов оптимизации. А пока еще немного подумаем о том, как можно закодировать данные в квантовую схему.

Кодирование данных в VQC

В конструировании VQC есть две важных части:

- кодирование классических данных в квантовые операторы;
- выбор наблюдаемой для измерений.

Вот как можно закодировать данные.

Кодирование поворотами

Один из самых популярных методов кодирования классических данных в квантовые схемы – это использование операторов поворота \hat{RY} , \hat{RZ} . Представим, что мы хотим отображать вектор из двумерного пространства в один кубит ($\vec{x} \in \mathbb{R}^2 \rightarrow \mathbb{C}^2$).

```

@qml.qnode(dev)
def angle_vqc(a, b):
    qml.RY(a, wires=[0])
    qml.RZ(b, wires=[0])

    return qml.expval(qml.PauliZ(0))

```

А теперь давайте посмотрим, как такая схема преобразует данные. Для этого сгенерируем набор случайных двумерных данных в диапазоне $[0, 2\pi]$ и применим к каждой точке нашей схемы, затем визуализируем результаты:

```

import numpy as np
import matplotlib.pyplot as plt

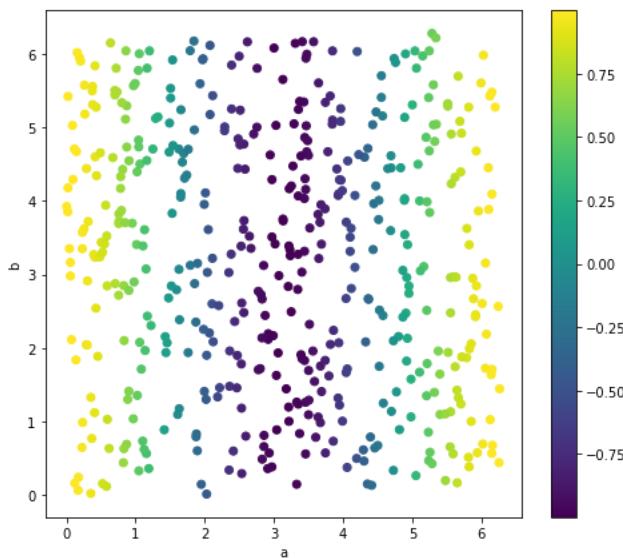
aa = np.random.uniform(0, 2 * np.pi, size=500)
bb = np.random.uniform(0, 2 * np.pi, size=500)

res = [angle_vqc(a, b) for a, b in zip(aa, bb)]

f = plt.figure(figsize=(8, 7))
clb = plt.scatter(aa, bb, c=res)
plt.xlabel("a")
plt.ylabel("b")
plt.colorbar(clb)

f.show()

```



Результат довольно необычный. Предсказать такое довольно сложно и хорошо видна возможность нелинейности VQC схем. В дальнейших лекциях мы будем более подробно разбирать возможные варианты кодирования данных. Кстати, кодирование углами настолько популярно, что в PennyLane для этого есть специальная функция, которая позволяет “поворнуть” сразу множество кубитов, используя список из классических данных:

```

@qml.qnode(dev)
def angle_vqc_pl(a, b):
    qml.templates.embeddings.AngleEmbedding([a], wires=[0], rotation="Y")
    qml.templates.embeddings.AngleEmbedding([b], wires=[0], rotation="Z")
    return qml.expval(qml.PauliZ(0))

```

Давайте убедимся, что это даст нам ровно тот же результат:

```

res_pl = [angle_vqc_pl(a, b) for a, b in zip(aa, bb)]
print(np.allclose(res, res_pl))

```

True

Важность многокубитных гейтов

При кодировании параметров целесообразно использовать многокубитные гейты, так как именно они создают запутанные состояния и раскрывают все преимущества квантовых компьютеров. Давайте попробуем это сделать.

```

dev2 = qml.device("default.qubit", 2)

@qml.qnode(dev2)
def cnot_vqc(a, b):
    qml.RY(a, wires=0)
    qml.RY(b, wires=1)
    qml.RZ(a, wires=0)
    qml.RZ(b, wires=1)
    qml.CZ(wires=[0, 1])
    return qml.expval(qml.PauliZ(0) @ qml.PauliZ(1))

```

В этой схеме мы видим оператор \hat{CZ} , который создает запутанное состояние. Посмотрим, как такая схема преобразует наше исходное пространство случайных точек.

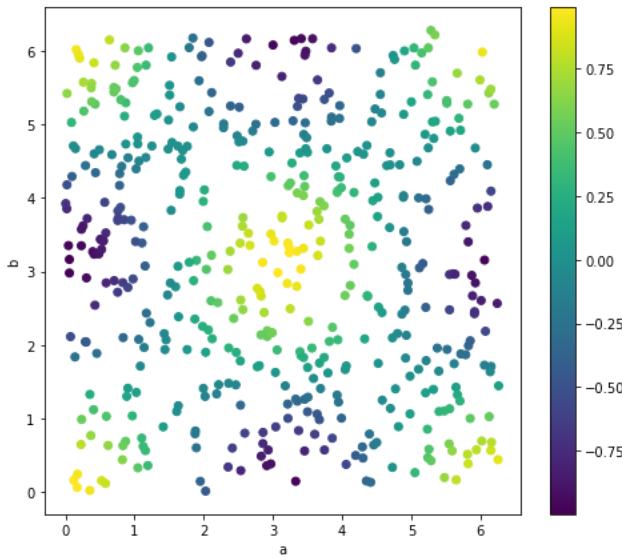
```

res = [cnot_vqc(a, b) for a, b in zip(aa, bb)]

f = plt.figure(figsize=(8, 7))
clb = plt.scatter(aa, bb, c=res)
plt.xlabel("a")
plt.ylabel("b")
plt.colorbar(clb)

plt.show()

```



Результат выглядит гораздо интереснее. В целом можно сказать, что выбор способа кодирования данных – один из важнейших при проектировании VQC. Мы еще много раз будем обсуждать способы сделать это.

Что мы узнали из лекции

- В эпоху NISQ мы чаще всего ограничены комбинированием квантового и классического машинного обучения.
- Ключевой элемент такого обучения – вариационные квантовые схемы (VQC). Основная идея VQC:
 - кодируем классические данные в квантовые операторы;
 - измеряем состояние;
 - варьируем параметры на классическом компьютере так, чтобы измерение давало желаемый результат.
- Один из широко применимых методов кодирования – использование операторов вращений \hat{RX} , \hat{RY} , \hat{RZ}

О блоке “Квантовые градиенты”

Этот блок включает в себя обзор способов оценки градиента VQC.

Продвинутая тема блока рассказывает о квантовых градиентах старших порядков.

Градиенты квантовых схем

Описание лекции

В этой лекции мы детально разберем, как можно оптимизировать параметры VQC:

- Как выглядит цикл обучения квантовой схемы
- Как работает оценка градиента "под капотом"
 - Метод конечных отрезков
 - Parameter-shift rule

Введение

Как мы уже говорили ранее, VQC выступают в роли "черных ящиков", которые имеют параметры и как-то преобразуют поступающие в них данные. В этом случае сам процесс оптимизации параметров выполняется на классическом компьютере. Одними из самых эффективных на сегодня методов решения задач непрерывной оптимизации являются градиентные методы. Для этих методов разработан широкий арсенал эвристик и приемов, который применяется в обучении классических глубоких нейронных сетей. Очень хочется применить весь этот арсенал и для квантового машинного обучения. Но как же посчитать градиент вариационной квантовой схемы?

Задача лекции

На этой лекции мы рассмотрим простую задачку по оптимизации параметров квантовой схемы и на ее примере увидим, как работают квантовые градиенты. В качестве задачи возьмем известный набор данных "Two Moon" из библиотеки `scikit-learn`:

```
from pennylane import numpy as np
import pennylane as qml
import matplotlib.pyplot as plt
%config InlineBackend.figure_format = 'retina'
from sklearn.datasets import make_moons

x, y = make_moons(n_samples=50)
```

Для удобства мы сразу переведем метки классов из $\{0, 1\}$ в $\{-1, 1\}$, а признаки \mathbf{X} переведем в $[0, \pi]$:

```
def normalize(x):
    """
    Переводит значения в интервал от 0 до pi
    """
    min_ = x.min()
    max_ = x.max()
    return np.pi * (x - min_) / (max_ - min_)

x[:, 0] = normalize(x[:, 0])
x[:, 1] = normalize(x[:, 1])
y = y * 2 - 1

plt.figure(figsize=(4, 3))
cb = plt.scatter(x[:, 0], x[:, 1], c=y)
plt.colorbar(cb)
plt.show()
```



Вариационная схема

Перед тем как мы начнем обучение и будем считать градиенты, нам необходимо определиться с тем, как будет выглядеть наша вариационная схема. Мы посвятим кодированию данных, выбору архитектуры схемы, а также измеряемого оператора еще много занятий. Так что пока просто воспользуемся кодированием признаков $\langle \mathbf{X} \rangle$ вращениями, а сверху применим несколько параметризованных слоев вращений.

Кодирование признаков

```
dev = qml.device("default.qubit", 2)

def encoding(x1, x2):
    qml.RY(x1, wires=0)
    qml.RY(x2, wires=1)
    qml.RZ(x1, wires=0)
    qml.RZ(x2, wires=1)
    qml.CZ(wires=[0, 1])
```

Параметризованные слои

В качестве одного слоя обучения мы будем использовать параметризованные вращения в связке с двухкубитным гейтом для создания запутанных состояний.

Note

Более детально о запутанных состояниях, а также квантовой энтропии и черных дырах можно посмотреть в [продвинутой лекции](#) блока про квантовые вычисления.

В этой лекции у нас нет цели идеально решить поставленную задачу – на самом деле это чуть сложнее, чем может показаться на первый взгляд. Поэтому пока не будем излишне усложнять нашу **VQC**. Сделаем нашу **VQC** содержащей несколько “слоев” следующего вида:

- Вращение 1-го кубита $\langle \hat{\text{Rot}}(\theta_1, \theta_2, \theta_3) \rangle$
- Вращение 2-го кубита $\langle \hat{\text{Rot}}(\theta_4, \theta_5, \theta_6) \rangle$
- “Запутывающий” оператор, который действует на оба кубита сразу – в нашем случае это $\langle \hat{\text{CZ}} \rangle$

Как видно, на каждый “слой” у нас приходится шесть параметров. Реализуем это в коде:

```
def layer(theta):
    qml.Rot(theta[0, 0], theta[0, 1], theta[0, 2], wires=0)
    qml.Rot(theta[1, 0], theta[1, 1], theta[1, 2], wires=1)
    qml.CZ(wires=[0, 1])
```

Здесь у нас вращения каждого из кубитов по сфере Блоха и двухкубитное взаимодействие $\langle \hat{\text{CZ}} \rangle$.

Теперь давайте объединим все это вместе, добавим пару наблюдаемых и оформим как `qml.qnode`:

```
@qml.qnode(dev)
def node(x1, x2, q):
    encoding(x1, x2)
    for q_ in q:
        layer(q_)

    return qml.expval(qml.PauliZ(0) @ qml.PauliY(1))
```

Функция “скоринга”

Наша квантовая схема принимает на вход лишь одну точку данных, а у нас их 50. Поэтому удобно сразу написать функцию, которая может работать с массивами NumPy:

```
def apply_node(x, q):
    res = []

    for x_ in x:
        vqc_output = node(x_[0], x_[1], q[0])
        res.append(vqc_output + q[1])

    return res
```

Может показаться немного запутанно, но так получилось. Дело в том, что параметры схемы это только углы поворотов. Но мы также хотим добавить еще и смещение, поэтому `tuple` параметров у нас содержит два элемента: массив параметров схемы, а также значение смещения. Так как схема у нас принимает на вход лишь одну пару значений (x_1, x_2) , то для того, чтобы “проскорить” массив данных мы должны:

- итерироваться по строкам двумерного массива
- для каждой строки вычислять результат схемы – это функция от (x_1, x_2, θ) – массив параметров θ у нас первый элемент `tuple`
- добавлять смещение – это второй элемент `tuple`
- результат добавлять в итоговый массив

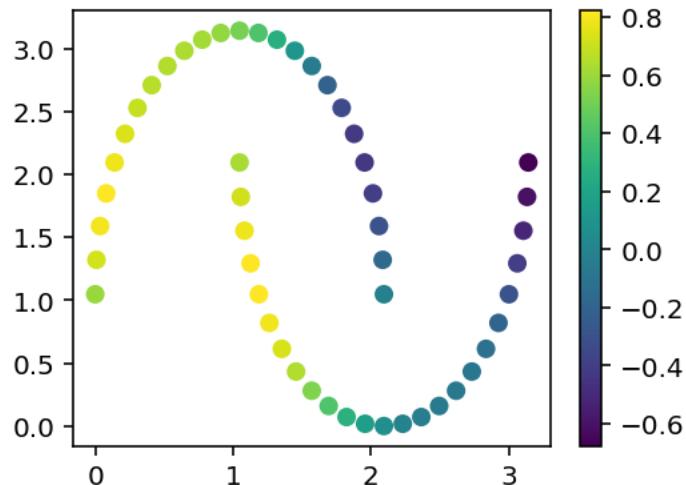
Именно это и реализовано в коде.

Визуализация

Давайте инициализируем нашу схему случайными параметрами и посмотрим, как она “сходу” классифицирует данные. Возьмем 4 параметризованных слоя.

```
np.random.seed(42)
q = (np.random.uniform(-np.pi, np.pi, size=(4, 2, 3)), 0.0)

plt.figure(figsize=(4, 3))
cb = plt.scatter(x[:, 0], x[:, 1], c=apply_node(x, q))
plt.colorbar(cb)
plt.show()
```



Как видно, результат “не очень” и наша цель – попытаться его улучшить.

Функция потерь

Прежде чем варьировать параметры схемы, нам для начала необходимо понять, а что именно мы хотим оптимизировать. Для этого нам необходимо выбрать функцию потерь.

💡 Note

Если у Вас трудности с функциями потерь в таком контексте, то рекомендуем вернуться к вводной лекции про классическое машинное обучение, где эта тема раскрыта достаточно подробно.

Квадратичное отклонение

В качестве функции потерь, которая является дифференцируемой, мы будем использовать наиболее простой вариант – среднеквадратичное отклонение. Это не самый лучший выбор для задач классификации, но зато самый простой. Простой вариант – это именно то, что нам нужно в этой лекции:

```
def cost(q, x, y):
    preds = np.array(apply_node(x, q))
    return np.mean(np.square(preds - y))
```

Точность классификации

В качестве метрики качества среднеквадратичное отклонение вообще не подходит – понять по этой цифре, хорошо или плохо работает модель почти невозможно! Поэтому для оценки модели в целом мы будем использовать точность:

```
def acc(q, x, y):
    preds = np.sign(apply_node(x, q))
    res = 0
    for p_, y_ in zip(preds, y):
        if np.abs(y_ - p_) <= 1e-2:
            res += 1
    return res / y.shape[0]
```

Решение средствами PennyLane

Библиотека [PennyLane](#) может использовать один из нескольких движков для автоматического дифференцирования:

- [NumPy Autograd](#)
- [PyTorch](#)
- [Tensorflow](#)
- [Jax](#)

По больше части, на наших занятиях мы будем использовать [NumPy](#) из-за простоты и привычности. Перед тем как разбираться с тем, как же именно происходит дифференцирование квантовой схемы, давайте посмотрим на весь цикл обучения.

💡 Note

Внимание, процесс обучения на обычном ноутбуке может занять около минуты! Это связано с трудностью симуляции квантового компьютера на классическом.

```

opt = qml.optimize.GradientDescentOptimizer(stepsizes=0.05)
acc_ = []
cost_ = []
ii = []
for i in range(75):
    batch = np.random.randint(0, len(x), (10,))
    x_batch = x[batch, :]
    y_batch = y[batch]
    q = opt.step(lambda q_: cost(q_, x_batch, y_batch), q)

    if i % 5 == 0:
        ii.append(i)
        acc_.append(acc(q, x, y))
        cost_.append(cost(q, x, y))

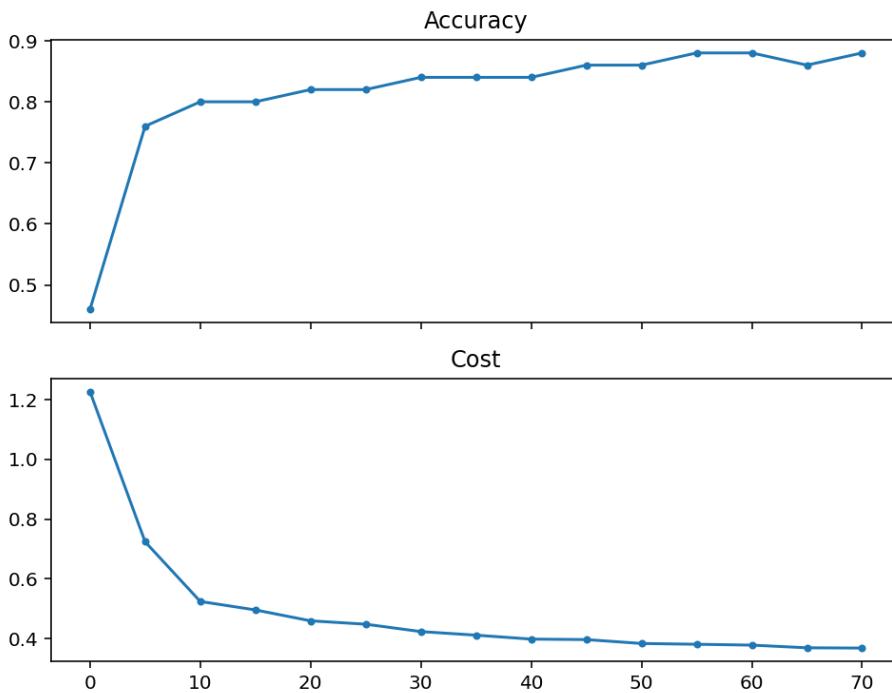
```

И посмотрим на получившиеся графики точности и функции потерь:

```

f, ax = plt.subplots(2, figsize=(8, 6), sharex=True)
ax[0].plot(ii, acc_, ".-")
ax[0].set_title("Accuracy")
ax[1].plot(ii, cost_, ".-")
ax[1].set_title("Cost")
plt.show()

```



А также на результаты классификации:

```

plt.figure(figsize=(4, 3))
cb = plt.scatter(x[:, 0], x[:, 1], c=apply_node(x, q))
plt.colorbar(cb)
plt.show()

```



А как оно работает?

Теперь, когда мы увидели процесс оптимизации квантовой схемы, давайте попробуем подумать, а как оно на самом деле работает?

Метод конечных отрезков

Для начала вспомним то, что является геометрическим (или визуальной) интерпретацией градиента функции. Правильно, градиент в каждой точке – это касательная. А приближенное значение угла наклона любой прямой можно найти, взяв конечные отрезки:

$$\frac{df}{dx} \approx \frac{\Delta f}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

Давайте попробуем посчитать градиент нашей квантовой схемы. Для этого инициализируем ее заново случайными параметрами, а потом сравним это с тем значением, которое считает [Autograd](#).

```
np.random.seed(42)
q = (np.random.uniform(0, np.pi, size=(4, 2, 3)), 0.0)

def naive_grad(cost, params, x, y, f0, delta):
    return (cost(params, x, y) - f0) / delta

def grad_i(q, f0, cost, x, y, i):
    new_params = q[0].copy().flatten()
    new_params[i] += 0.05

    return naive_grad(cost, (new_params.reshape(q[0].shape), q[1]), x, y, f0, 0.05)
```

В качестве эталона возьмем тот градиент, который нам считает [Autograd](#):

```
grad =qml.grad(cost, argnum=0)
```

Сравним первые пять значений [Autograd](#) с нашим наивным алгоритмом, взяв $(\Delta x = 0.05)$:

Note

Осторожно! Так как мы считаем градиенты очень наивно и на всех точках сразу, то следующий блок кода работает долго!

```
autograd = grad(q, x, y)
f0 = cost(q, x, y)
pretty_print = ""
for i in range(10):
    g_i = grad_i(q, f0, cost, x, y, i)
    pretty_print += f"Naive grad: {g_i:.3f}\tAutograd result: {autograd[0].flatten()}"
    [i]:.3f\n"
print(pretty_print)
```

```

Naive grad: -0.028      Autograd result: -0.029
Naive grad: 0.081       Autograd result: 0.081
Naive grad: 0.025       Autograd result: 0.025
Naive grad: 0.013       Autograd result: 0.008
Naive grad: -0.000      Autograd result: -0.011
Naive grad: 0.009       Autograd result: 0.004
Naive grad: 0.025       Autograd result: 0.025
Naive grad: 0.002       Autograd result: -0.003
Naive grad: -0.068      Autograd result: -0.069
Naive grad: 0.009       Autograd result: 0.004

```

Можно заметить, что даже с таким большим значением $\|\Delta\|$ наши оценки получились достаточно близкими к тем, которые получены в **Autograd**. Хотя, конечно, для некоторых значений расхождения заметны и иногда они даже в знаке частной производной.

Parameter-shift rule

Более точная оценка может быть получена методом, который называется **Parameter-shift**. Он основан на том, что для квантового “черного ящика” $\langle \hat{U}(\theta) \rangle$, которым является наша схема, частная производная по параметру θ_i выражается так:

$$\langle \nabla_{\theta_i} \hat{U} \rangle = \frac{1}{2} (\langle \hat{U}(\theta + \pi/2) \rangle - \langle \hat{U}(\theta - \pi/2) \rangle)$$

Note

Более строгую формулировку, а также вывод правила parameter-shift можно посмотреть в продвинутой лекции этого блока про производные высших порядков.

Если по-простому, то оценка частной производной по θ_i -му параметру может быть получена вычислением сначала ожидаемого значения схемы с параметром θ_i , смещенным на $\pi/2$ в одну сторону, а потом – в другую. Давайте запишем это в коде, но перед этим давайте вспомним, как будет выглядеть производная функции потерь (а именно она нам нужна):

$$\frac{\partial L}{\partial \theta_i} = \frac{1}{2} (\langle \hat{U}(\theta + \pi/2) \rangle - \langle \hat{U}(\theta - \pi/2) \rangle)$$

Реализуем явно и наивно эту формулу в коде:

```

def parameter_shift_i(q, cost, x, y, i, y_hat):
    new_params = q[0].copy().flatten()
    new_params[i] += np.pi / 2

    forward = np.array(apply_node(x, (new_params.reshape(q[0].shape), q[1])))

    new_params = q[0].copy().flatten()
    new_params[i] -= np.pi / 2

    backward = np.array(apply_node(x, (new_params.reshape(q[0].shape), q[1])))

    diff = (y_hat - y)

    return np.mean(2 * diff * (0.5 * (forward - backward)))

```

И также проверим на первых 10 точках:

```

y_hat = apply_node(x, q)
pretty_print = ""
for i in range(10):
    g_i = parameter_shift_i(q, cost, x, y, i, y_hat)
    pretty_print += f"Naive grad: {g_i:.3f}\tAutograd result: {autograd[0].flatten()[i]:.3f}\n"
print(pretty_print)

```

Naive grad: -0.029	Autograd result: -0.029
Naive grad: 0.081	Autograd result: 0.081
Naive grad: 0.025	Autograd result: 0.025
Naive grad: 0.008	Autograd result: 0.008
Naive grad: -0.011	Autograd result: -0.011
Naive grad: 0.004	Autograd result: 0.004
Naive grad: 0.025	Autograd result: 0.025
Naive grad: -0.003	Autograd result: -0.003
Naive grad: -0.069	Autograd result: -0.069
Naive grad: 0.004	Autograd result: 0.004

Как видно, этот результат уже совпадает с тем, что делает “под капотом” [PennyLane](#) и [Autograd](#). На самом деле, правило **parameter-shift** позволяет использовать много интересных хитростей и оптимизаций, но их не получится легко показать без погружения в математические детали метода.

Что мы узнали?

- Мы попробовали провести полный цикл оптимизации параметров **VQC**
- Научились использовать автоматический расчет градиентов в [PennyLane](#)
- Познакомились с двумя способами оценки градиента:
 - Метод конечных отрезков
 - Parameter-shift rule

Градиенты высших порядков

План лекции

В этой лекции мы посмотрим на ту математику, которая лежит “под капотом” у *parameter-shift rule*. Мы познакомимся с обобщением *parameter shift*, а также увидим, как можно оптимизировать этот метод. В конце мы узнаем, как можно посчитать производную второго порядка за минимальное количество обращений к квантовому компьютеру.

Для более детального погружения в вопрос можно сразу рекомендовать статью [[MBK21](#)].

Важность гейтов вращений

Если задуматься, то одним из основных (если не единственных) способов сделать параметризованную квантовую схему является использование гейтов вращений, таких как $\{\hat{RX}, \hat{RY}, \hat{RZ}\}$. Более формально это можно выразить так, что нас больше всего интересуют операторы вида:

$$U(\theta) = e^{-\frac{i}{2}H\theta}$$

где H – оператор “вращения”, который удовлетворяет условию $H^2 = \mathbf{1}$. Другой возможный вариант записи – представить матрицу H как линейную комбинацию операторов Паули $(\sigma_x, \sigma_y, \sigma_z)$.

Если представить схему, содержащую множество параметризованных операторов, то итоговая запись имеет вид:

$$U_{[j...k]} = U_j, \dots, U_k | \Psi \rangle$$

Производная от измерения

Давайте вспомним, как выглядит квантово-классическая схема обучения с **VQC**.



Fig. 101 Квантово-классическая схема

Видно, что мы хотим считать производную не от самой параметризованной схемы $(U_{j\dots k})$, а от наблюдаемой. Для тех, кто забыл, что такое наблюдаемая, рекомендуем вернуться к [лекции про кубит](#). Если кратко, то это тот оператор, который мы “измеряем” на нашем квантовом компьютере. Математически производная, которая нам интересна, может быть записана для выбранного параметра (i) таким образом:

$$\langle G_i = \frac{\partial}{\partial \theta_i} \langle \Psi | \hat{M} | \Psi \rangle \rangle$$

То есть нам важно посчитать производную от результата измерения, так как именно результат измерения у нас будет определять “предсказание” нашей квантовой нейронной сети. Причем нам нужно уметь считать производную от любого параметра (θ_i) в цепочке $(\theta_j, \dots, \theta_i, \dots, \theta_k)$.

Parameter-shift для гейтов Паули

Note

Тут мы для простоты предложим, что (U_1) это просто оператор вращения, иначе выкладки станут совсем сложными.

Тогда сам оператор (U_i) может быть также записан так:

$$U_i = e^{-\frac{i}{2}\theta_i P_i}$$

Запишем результат математического ожидания через состояние $(|\Psi_i\rangle)$, которое пришло на вход (i) -го гейта в нашей последовательности:

$$\langle M(\theta_i) \rangle = \text{Tr}(M U_{\{k, \dots, 1\}} \rho_i U_{\{k, \dots, 1\}}^\dagger)$$

где (ρ_i) это матрица плотности $(\langle \Psi_i | \Psi_i \rangle)$. Подробнее о матрицах плотности можно почитать в ранней продвинутой лекции про смешанные состояния.

Тогда частная производная от математического ожидания по (i) -му параметру (θ_i) записывается (подробнее в [\[MNKF18\]](#)) через коммутатор исходного состояния (ρ_i) , которое “пришло” на вход гейта (U_i) и того оператора Паули (P_i) , который мы используем в (U_i) :

$$\frac{\partial}{\partial \theta_i} \langle M \rangle = -\frac{1}{2} \text{Tr}(M U_{\{k, \dots, i\}} [P_i, U_{\{i-1, \dots, 1\}}] \rho_i U_{\{i-1, \dots, 1\}}^\dagger)$$

Этот коммутатор может быть переписан следующим образом:

$$\begin{aligned} \langle [P_i, \rho_i] \rangle &= i[U_i, \text{Bigl}(\frac{\partial}{\partial \theta_i} \langle M \rangle \text{Bigr})] \rho_i U_i^\dagger - U_i \text{Bigl}(\frac{\partial}{\partial \theta_i} \langle M \rangle \text{Bigr}) \rho_i U_i^\dagger \end{aligned}$$

Тогда соответствующий градиент $(\frac{\partial}{\partial \theta_i} \langle M \rangle)$ можно записать через смещения на $(\pm \frac{\partial}{\partial \theta_i} \langle M \rangle)$:

$$\begin{aligned} \langle M \rangle &= \langle M \rangle^+ + \langle M \rangle^- \\ \langle M \rangle^+ &= \frac{1}{2} \text{Tr}[M U_{\{k, \dots, i+1\}} U_i^\dagger (\rho_i^+ + \rho_i^-)] \\ \langle M \rangle^- &= \frac{1}{2} \text{Tr}[M U_{\{k, \dots, i+1\}} U_i^\dagger (\rho_i^+ - \rho_i^-)] \end{aligned}$$

По аналогии с классическими нейронными сетями и *backpropagation* (для тех, кто забыл это понятие, рекомендуем вернуться к вводным лекциям про классическое машинное обучение) тут явно можно выделить *forward* проход со смещением (θ_i) на значения $(\frac{\partial}{\partial \theta_i} \langle M \rangle)$ и *backward* со смещением на $(-\frac{\partial}{\partial \theta_i} \langle M \rangle)$.

Обобщенный parameter-shift

Предложенное в [MNKF18] выражение может быть на самом деле получено в более общем виде из других соображений. Так, выражение для нашей наблюдаемой $\langle \hat{M} \rangle$ может всегда быть представлено [MBK21] как сумма вида:

$$\langle \hat{U}_i(\theta_i) \hat{M} \hat{U}_i^\dagger(\theta_i) \rangle = \hat{A} + \hat{B} \cos(\theta_i) + \hat{C} \sin(\theta_i)$$
 где $\hat{A}, \hat{B}, \hat{C}$ – операторы, не зависящие от параметра θ_i .

Note

Действительно, явно выписав выражение для наблюдаемой и вспомнив формулы для косинуса и синуса двойного угла, а также воспользовавшись тем, что $U(\theta) = e^{-\frac{i}{2}\theta \hat{H}}$, получаем:

$$\begin{aligned} & \langle \cos(\frac{\theta}{2}) \hat{M} \cos(\frac{\theta}{2}) \rangle = \cos^2(\frac{\theta}{2}) \hat{M} + i \sin(\frac{\theta}{2}) \cos(\frac{\theta}{2}) \hat{M} - i \sin(\frac{\theta}{2}) \cos(\frac{\theta}{2}) \hat{M} + i \sin^2(\frac{\theta}{2}) \hat{M} = \\ & \frac{1}{2} \left(\hat{M} + i \sin(\theta) \hat{M} + i \sin(\theta) \hat{M} - i \sin^2(\theta) \hat{M} \right) = \frac{1}{2} \left(\hat{M} + 2i \sin(\theta) \hat{M} - \sin^2(\theta) \hat{M} \right) = \frac{1}{2} \left(\hat{M} + 2i \sin(\theta) \hat{M} - (1 - \cos^2(\theta)) \hat{M} \right) = \end{aligned}$$

Тогда можно воспользоваться правилами тригонометрии, а именно, тем что для любого $s \neq k\pi$, $\text{atan}(s) \in \{1, 2, \dots, N\}$ справедливо:

$$\langle \frac{d \cos(\theta)}{d \theta} \rangle = \frac{\cos(\theta + s) - \cos(\theta - s)}{2s} = \frac{\sin(s)}{2s}$$

И подставим это в выражение для $\langle \partial_{\theta_i} \hat{M} \rangle$:

$$\langle \frac{\partial \hat{M}}{\partial \theta_i} \rangle = \frac{\langle \hat{M}(\theta_i + s) - \hat{M}(\theta_i - s) \rangle}{2s}$$

Легко заметить, что подстановка сюда $s = \frac{\pi}{2}$ дает нам классический *parameter shift*, описанный в [MNKF18].

Наконец, запишем полученное выражение в более удобном виде, который позволит нам более эффективно выписывать производные высших порядков. Для этого введем вектор \mathbf{e}_i – единичный вектор для i -го параметра, то есть вектор, где все компоненты кроме i -й равны нулю, а i -я равна 1. Тогда наше финальное выражение для обобщенного *parameter shift* примет следующий вид:

$$\boxed{\frac{\partial f(\hat{M}(\theta))}{\partial \theta_i} = \frac{f(\hat{M}(\theta + s\mathbf{e}_i)) - f(\hat{M}(\theta - s\mathbf{e}_i))}{2s}}$$

Вторая производная и гессиан

В классической теории оптимизации, также как и в машинном обучении, очень часто на первый план выходят так называемые методы 2-го порядка. Эти методы похожи на обычный градиентный спуск, но для ускорения сходимости они также используют информацию из матрицы вторых производных, которая называется гессианом. Более подробно про методы 2-го порядка и гессиан можно посмотреть в вводных лекциях курса.

Методы второго порядка требуют больше вызовов, чтобы вычислить гессиан, но взамен они обеспечивают гораздо лучшую сходимость, а также менее склонны “застревать” в локальных минимумах. Это обеспечивает, в итоге, более быстрое обучение. В классических нейронных сетях вычисление гессиана это часто проблема, так как это матрица размерности $O(N^2)$, где N – число весов нейронной сети, и эта матрица получается слишком большой. Но, как мы помним, основная “фича” VQC это их экспоненциальная экспрессивность – возможность линейным числом параметров (и гейтов) обеспечить преобразование, эквивалентное экспоненциальному числу весов классической нейронной сети. А значит, для них проблема размерности гессиана не стоит так остро. При этом использование гессиана теоретически позволяет в итоге обучить VQC за меньшее число вызовов. Именно поэтому методы второго порядка потенциально очень интересны в квантово-классическом обучении. Но для начала нам необходимо разобраться, как именно можно посчитать матрицу вторых производных.

Пользуясь обобщенным правилом *parameter shift*, можно выписать выражение для второй производной [MBK21]:

$$\begin{aligned} \frac{\partial^2 f}{\partial \theta_i^2} &= \frac{f(\hat{M}(\theta + s_1\mathbf{e}_i) + s_2\mathbf{e}_j) - f(\hat{M}(\theta + s_1\mathbf{e}_i) - s_2\mathbf{e}_j) - f(\hat{M}(\theta - s_1\mathbf{e}_i) + s_2\mathbf{e}_j) + f(\hat{M}(\theta - s_1\mathbf{e}_i) - s_2\mathbf{e}_j)}{4s_1s_2} \end{aligned}$$

Взяв $s_1 = s_2$, можно упростить это выражение к следующему виду:

```
\begin{gathered}
\frac{f(\mathbf{theta}) + s(\mathbf{a}) + f(\mathbf{theta}) + s(\mathbf{b}) - f(\mathbf{theta}) + s(\mathbf{c}) - f(\mathbf{theta}) + s(\mathbf{d})}{2 \sin(s)^2} \\
\mathbf{a} = \mathbf{e_i} + \mathbf{e_j} \\
\mathbf{b} = -\mathbf{e_i} - \mathbf{e_j} \\
\mathbf{c} = \mathbf{e_i} - \mathbf{e_j} \\
\mathbf{d} = -\mathbf{e_i} + \mathbf{e_j}
\end{gathered}
```

Но чаще всего нам необходимо не просто посчитать гессиан, а еще и посчитать градиент, так как в большинстве методов 2-го порядка требуются оба эти значения. В этом случае хочется попробовать подобрать такое значение для $\backslash(s_g)$ при вычислении вектора градиента, а также такое значение $\backslash(s_h)$ при вычислении гессиана, чтобы максимально переиспользовать результаты квантовых вызовов и уменьшить их общее количество.

Внимательно взглянув на выражение для 2-х производных, можно заметить, что оптимизация там возможна при расчете диагональных элементов гессиана. Давайте выпишем выражение для диагонального элемента явно:

```

\begin{array}{l}
\frac{f(\theta) + 2s e_i)}{f(\theta) - 2s e_i)} \cdot \frac{(2 \sin s)^2}{} \\

```

Можно заметить, что, например, использование $\backslash s = \frac{\pi}{4}$ для гессиана, а также "стандартного" $\backslash s = \frac{\pi}{2}$ для градиента позволит полностью переиспользовать в диагональных элементах гессиана значения, которые

мы получили при расчете градиента. А значение $\{f(\mathbf{\theta})\}$ вообще считается один раз для всех диагональных вызовов.

 Note

На самом деле, диагональные элементы гессиана можно использовать и сами по себе, например для квазиньютоновских методов оптимизации, где матрица Гессе аппроксимируется какой-то другой матрицей, чтобы не считать все вторые производные. Например, она может быть аппроксимирована диагональной матрицей, как в работе [And19].

Заключение

В этой лекции мы познакомились с классическим *parameter shift rule*, а также его обобщением. Также мы узнали, как можно посчитать гессиан **VQC**, и даже узнали маленькие хитрости, которые можно применять для уменьшения общего количества вызовов квантовой схемы.

О блоке “Квантовый SVM”

Этот блок включает в себя только квантово-классический SVM.

Продвинутая тема блока дополнительно рассказывает о классическом SVM.

Квантово-классический SVM

Описание лекции

Лекция будет построена следующим образом:

- посмотрим, как можно использовать **VQC** как ядро SVM;
 - напишем и применим код обучения смешанного SVM.

Перед тем как приступить, рекомендуется освежить [материал](#) по классическому методу опорных векторов, в частности, нужно понимать, как работает ядерный трюк.

VQC как ядерная функция

Как мы много говорили в более ранних лекциях, квантовые схемы позволяют нам оперировать в гильбертовых пространствах волновых функций. Эти пространства имеют экспоненциально большую размерность, при этом они параметризуются линейным количеством параметров. А еще в этих пространствах определены скалярные произведения волновых функций, более того, именно результат выборки из скалярного произведения мы чаще всего и получаем как результат измерений!

Давайте попробуем посмотреть, что общего у **VOC** и ядер.

- оба оперируют в пространстве большой (или бесконечной размерности);
 - и там, и там работа идет в гильбертовом пространстве и определено скалярное произведение;
 - и там, и там результат вычисляется как скалярное произведение.

Note

Это интересно, что многие специалисты в области **QML** сегодня даже предлагают вместо термина “квантовая нейросеть” использовать термин “квантовое ядро”, так как математически **VQC** гораздо ближе именно к ядрам, чем к слоям современных глубоких сетей.

Очевидная идея – попробовать как-то воспользоваться квантовой схемой, чтобы реализовать скалярное произведение двух классических векторов. Именно это и сделали авторы работы [[HaylivcekCorcolesT+19](#)].

Преобразование состояния

На самом деле, если просто использовать какие-то простые квантовые операции, мы не получим какого-то преимущества над классическим ядерным SVM – ведь все то же самое можно будет сделать и на классическом компьютере.

Чтобы получить реальное преимущество, нам необходимо использовать запутывание и прочие “фишки” квантовых вычислений.

Дальше мы не станем изобретать велосипеды, а вместо этого воспользуемся примерами хороших преобразований из работы [[SYG+20](#)]. Рассмотрим, что именно там описано.

Общая схема

Для простоты формул мы не будем выписывать обобщенные формулы, а все будем писать для нашего двумерного пространства. Тогда наша схема может быть разделена на несколько частей:

- гейты Адамара и гейты $\backslash(CNOT)$;
- операции, основанные на элементах входного вектора;
- попарные операции над парами элементов вектора.

Мы начинаем с того, что переводим кубиты в состояние суперпозиции, применяя операторы Адамара. Далее мы применяем однокубитные параметризованные операции и снова гейты Адамара. После этого мы применяем связку $\backslash(CNOT \backslash to)$ параметризованная парой операция $\backslash(to CNOT)$.

Выбор операции

Следуя идее упомянутой статьи, в качестве что одно-элементной, что двух-элементной операции мы будем использовать гейт $\backslash(U_1)$. Разница будет лишь в том, что мы передаем на вход в качестве параметра.

Feature function

В качестве параметров на входе гейта $\backslash(U_1)$, как мы уже говорили, выступают один или два элемента вектора $\backslash(x)$. Строго это можно записать как функцию такого вида:

$$\begin{aligned} \backslash\begin{cases} \text{split} \\ \phi(x_1, x_2) = \begin{cases} \phi(x), & \text{if } x_1 = x_2 \\ \phi(x_1, x_2), & \text{if } x_1 \neq x_2 \end{cases} \end{cases} \end{aligned}$$

Мы будем называть ее *feature function*. В некотором смысле можно сказать, что именно эта функция определяет тип ядра по аналогии с классическим SVM. В работе [[SYG+20](#)] описано много разных вариантов таких *feature function*, мы будем использовать следующую:

$$\begin{aligned} \backslash\begin{cases} \text{split} \\ \text{gathered} \\ \phi(x) = x \backslash\phi(x_1, x_2) = \pi \cos(x_1) \cos(x_2) \end{cases} \end{aligned}$$

Скалярное произведение

Все что мы описали выше, обозначим как квантовую схему $\backslash(U(x))$. Она преобразует нам вектор классических данных $\backslash(x)$ в квантовое состояние $\backslash(\ket{\Psi})$. Но нам то нужно получить скалярное произведение $\backslash(\bra{x}_1 \backslash U(x_2))$. Выглядит сложно, но на самом деле существует эффективный способ получить эту величину без необходимости восстанавливать весь вектор состояния. Можно показать, что величина $\backslash(\bra{x}_1 \backslash U(x_2))$ равна вероятности нулевой битовой строки ($\backslash(\ket{0, 0, \dots, 0})$) при измерении другой схемы: $\backslash(U(x_1)U(x_2)^{\dagger})$.

Все это может казаться сложным и запутанным, но должно стать гораздо понятнее, когда мы посмотрим на пример реализации от начала и до конца.

Пример реализации

Схема

Для начала необходимые импорты.

```
from pennylane import numpy as np
import pennylane as qml
import matplotlib.pyplot as plt
%config InlineBackend.figure_format = 'retina'
from sklearn.datasets import make_moons
```

Помимо всех привычных, нам еще потребуется классический SVM из `scikit-learn`:

```
from sklearn.svm import SVC
```

Мы будем работать с уже привычным нам набором "Two Moons". Только в этом случае мы будем использовать чуть-чуть другую нормализацию – для нашего ядра элементы вектора $\langle x \rangle$ должны быть в интервале $[-1, 1]$.

Сразу переведем наши данные в этот диапазон:

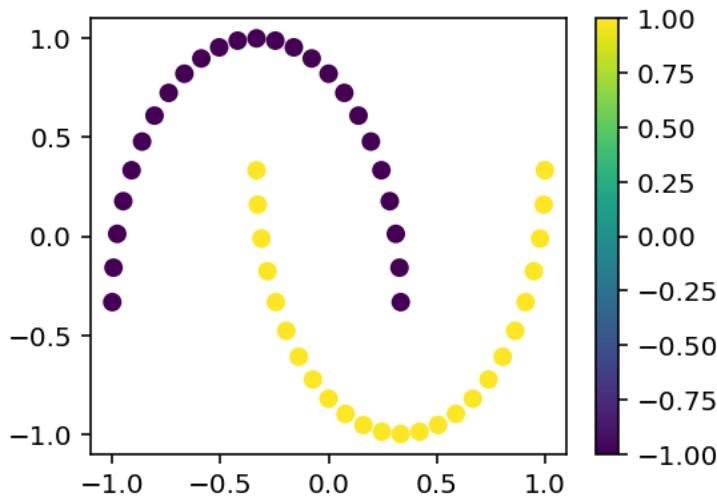
```
x, y = make_moons(n_samples=50)
y = y * 2 - 1

def normalize(x):
    """Переводит значения в интервал от -1 до 1"""

    min_ = x.min()
    max_ = x.max()
    return 2 * (x - min_) / (max_ - min_) - 1

x[:, 0] = normalize(x[:, 0])
x[:, 1] = normalize(x[:, 1])

plt.figure(figsize=(4, 3))
cb = plt.scatter(x[:, 0], x[:, 1], c=y)
plt.colorbar(cb)
plt.show()
```



И привычное нам объявление устройства.

```
dev = qml.device("default.qubit", 2)
```

Теперь давайте для начала реализуем наше преобразование над одним из векторов ($\langle U(x) \rangle$). Поскольку далее нам потребуется еще и $\langle U(x)^\dagger \rangle$, то мы сразу воспользуемся декоратором `@qml.template`, который позволит нам автоматически получить обратную схему.

```

@qml.template
def var_layer(x):
    qml.Hadamard(wires=0)
    qml.Hadamard(wires=1)

    qml.U1(x[0], wires=0)
    qml.U1(x[1], wires=1)

    qml.Hadamard(wires=0)
    qml.Hadamard(wires=1)

    qml.CNOT(wires=[0, 1])
    qml.U1(np.pi * np.cos(x[0]) * np.cos(x[1]), wires=1)
    qml.CNOT(wires=[0, 1])

```

А теперь реализуем

$$\langle \Psi | U(x_1) U(x_2) = \langle \Psi | U(x_1) U(x_2)^\dagger M_0 | \Psi \rangle$$

где M_0 – проектор на один из собственных векторов системы кубитов, а именно на “нулевой”:

$$|M_0\rangle = |\{0, \dots, 0\}\rangle$$

Проще говоря, мы реализуем схему, которая нам дает вероятности каждой из битовых строк (а дальше мы просто возьмем первую, она и отвечает строке $|0, \dots, 0\rangle$):

```

@qml.qnode(dev)
def dot_prod(x1, x2):
    var_layer(x1)
    qml.inv(var_layer(x2))

    return qml.probs(wires=[0, 1])

```

Ну и сразу вспомогательную функцию, которая нам считает то, что нам было нужно:

```

def q_dot_prod(i, j):
    x1 = (x[i, 0], x[i, 1])
    x2 = (x[j, 0], x[j, 1])
    return dot_prod(x1, x2)[0]

```

Для самопроверки убедимся в том, что наше “скалярное произведение” симметрично:

```
print(np.allclose(q_dot_prod(0, 1), q_dot_prod(1, 0)))
```

True

```
/home/runner/work/qmlcourse/qmlcourse/.venv/lib/python3.8/site-
packages/pennylane/utils.py:371: UserWarning: Use of qml.inv() is deprecated and
should be replaced with qml.adjoint().
warnings.warn
```

И сразу посмотрим на то, как выглядит наша схема:

```
print(dot_prod.draw())
```

```

0: ─H─Rφ(0.333)─H─rC────────rC─rC────────rC─H─Rφ(-0.246)─H─r|
Probs
1: ─H─Rφ(-0.333)─H─lX─Rφ(2.81)─lX─lX─Rφ(-1.68)─lX─H─Rφ(0.989)─H─l|
Probs

```

Гибридный SVM

Мы не будем сами с нуля писать решение задачи квадратичного программирования, просто воспользуемся готовой рутиной из [scikit-learn](#). Используемая там реализация позволяет вместо ядерной функции передать сразу матрицу Грама ([Gram matrix](#)). На самом деле это просто матрица всех попарных скалярных произведений наших векторов. Вычислим ее, сразу воспользовавшись тем, что

$$\langle U(x) | U(x) \rangle = 1 \text{ и } \langle U(x_1) | U(x_2) \rangle = \langle U(x_2) | U(x_1) \rangle$$

```

gram_mat = np.zeros((x.shape[0], x.shape[0]))

for i in range(x.shape[0]):
    for j in range(x.shape[0]):
        if i == j:
            gram_mat[i, j] = 1
        if i > j:
            r = q_dot_prod(i, j)
            gram_mat[i, j] = r
            gram_mat[j, i] = r

```

Обучим нашу модель:

```

model = SVC(kernel="precomputed")
model.fit(gram_mat, y)

```

```
SVC(kernel='precomputed')
```

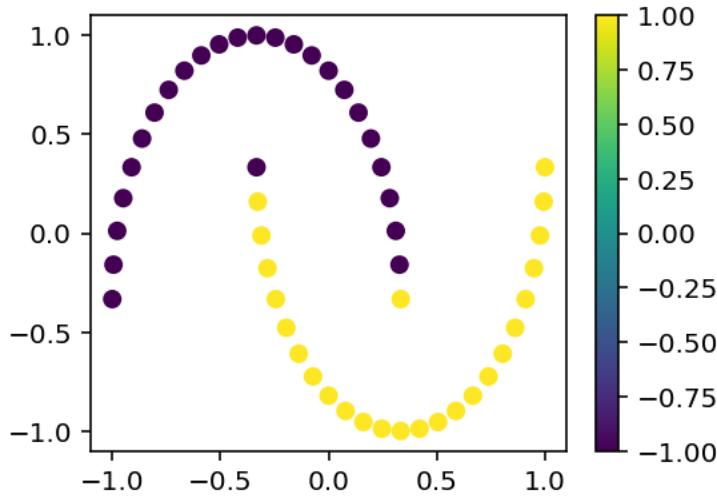
Посчитаем предсказания и посмотрим на результат:

```

preds = model.predict(X=gram_mat)

plt.figure(figsize=(4, 3))
cb = plt.scatter(x[:, 0], x[:, 1], c=preds)
plt.colorbar(cb)
plt.show()

```



Результат выглядит неплохо!

Заключение

Расчет полной матрицы скалярных произведений дает нам сложность $\mathcal{O}(N^2)$ вызовов. Но основной потенциал гибридного SVM в том, что задачу квадратичной оптимизации на самом деле можно тоже решать на квантовом компьютере, используя алгоритм Гровера (про него рассказано в ранних лекциях), причем за сложность всего $\mathcal{O}(N)$ и без расчета полной матрицы Грамма!

Многие считают, что NISQ квантовые компьютеры могут стать для SVM чем-то типа видеокарт для нейронных сетей и вернуть этот алгоритм на пьедестал лучших алгоритмов машинного обучения!

Классический SVM

Описание лекции

В этой лекции мы рассмотрим классический метод опорных векторов (SVM) и разберем стоящую за ним математику. С согласия Евгения Соколова, мы во многом переиспользуем конспекты лекций [курса](#) “Машинное обучение”, читаемого на ФКН ВШЭ.

План лекции такой:

- линейная классификация;
- интуиция метода опорных векторов;
- метод опорных векторов для линейно-разделимой выборки;
- метод опорных векторов для линейно-неразделимой выборки;
- решение задачи метода опорных векторов;
- ядерный переход;
- плюсы и минусы SVM.

Линейная классификация

Основная идея линейного классификатора заключается в том, что признаковое пространство может быть разделено гиперплоскостью на два полупространства, в каждом из которых прогнозируется одно из двух значений целевого класса. Если это можно сделать без ошибок, то обучающая выборка называется *линейно разделимой*.

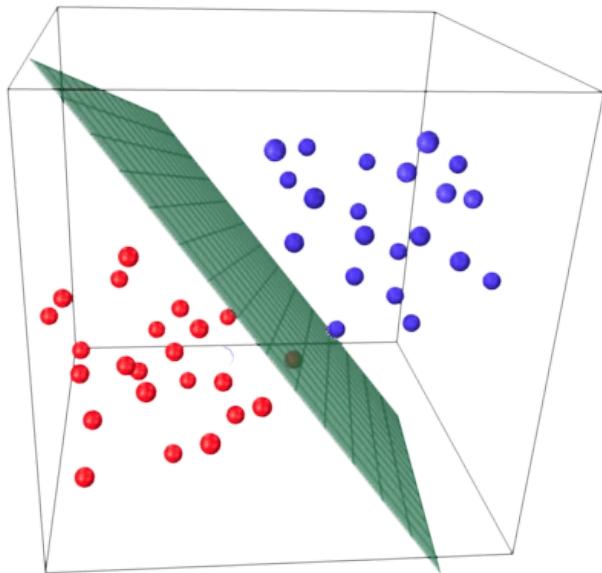


Fig. 102 Линейная классификация

Рассмотрим задачу бинарной классификации, в которой $\mathbb{X} = \mathbb{R}^d$ – пространство объектов, $\mathbb{Y} = \{-1, +1\}$ – множество допустимых ответов (целевой признак), $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^n$ – обучающая выборка. Будем класс $+1$ называть положительным, а класс -1 – отрицательным. Здесь d – размерность признакового пространства, n – количество примеров в обучающей выборке.

Линейная модель классификации определяется следующим образом:

$a(x) = \text{sign}(\langle w, x \rangle + b) = \text{sign}(\sum_{j=1}^d w_j x_j + b)$,
где $w \in \mathbb{R}^d$ – вектор весов, $b \in \mathbb{R}$ – сдвиг (bias), $\text{sign}(\cdot)$ – функция “сигнум”, возвращающая знак своего аргумента, $a(x)$ – ответ классификатора на примере x .

Note

Хорошо бы для векторов ставить стрелку и писать $\langle \vec{w} \rangle$, но мы этого не будем делать, предполагая, что из контекста ясно, что вектор, а что скаляр. В частности, в формуле выше w и x – вектора, а $a(x)$, w_j , x_j , d и b – скаляры.

Часто считают, что среди признаков есть константа, $x_{d+1} = 1$. В этом случае нет необходимости вводить сдвиг b , и линейный классификатор можно задавать как

$a(x) = \text{sign}(\langle w, x \rangle)$.

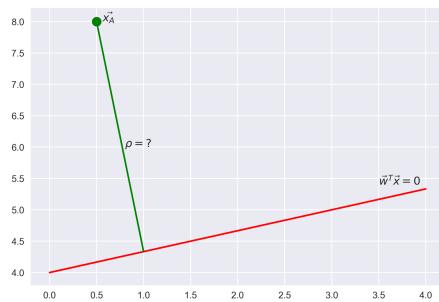


Fig. 103 Расстояние от точки до плоскости

Геометрически линейный классификатор соответствует гиперплоскости с вектором нормали $\|w\|$, которая задается уравнением $\langle w, x \rangle = 0$. Величина скалярного произведения $\langle w, x \rangle$ пропорциональна расстоянию от гиперплоскости до точки $|x|$, а его знак показывает, с какой стороны от гиперплоскости находится данная точка. Если быть точным, расстояние от точки с радиус-вектором $|x_A|$ до плоскости $\langle w, x \rangle = 0$:

$$|\rho(x_A, \langle w, x \rangle = 0)| = \frac{|\langle w, x_A \rangle|}{\|w\|}.$$

❶ Упражнение на метод Лагранжа

[Метод Лагранжа](#) – очень важный метод оптимизации функций при наличии ограничений. Этот же метод вовсю используется ниже в методе опорных векторов.

Рекомендуется ознакомиться с методом Лагранжа и решить (ручками!) простую оптимизационную задачу вида

$$\begin{aligned} & \left| \begin{aligned} & \text{min}_{x,y} & & x + y \\ & \text{subject to} & & x^2 + y^2 = 1 \end{aligned} \right. \\ & \end{aligned}$$

А мы выведем указанную выше формулу расстояния от точки до плоскости. Вообще это можно сделать разными способами – алгебраически и геометрически. Но давайте посмотрим на эту задачу (внезапно) как на задачу оптимизации и решим ее методом Лагранжа. Это послужит неплохой тренировкой перед тем как окунуться в SVM.

Итак, представим задачу в таком виде: хотим найти точку $|x|$ на плоскости $\langle w, x \rangle = 0$ такую, что расстояние от $|x|$ до точки $|x_0|$ минимально:

$$\left| \begin{aligned} & \text{min}_x & & |\rho(x, x_0)| \\ & \text{subject to} & & \langle w, x \rangle = 0 \end{aligned} \right.$$

Лагранжиан: $\mathcal{L}(x, \lambda) = \|x - x_0\|^2 + \lambda \langle w, x \rangle$. Тут мы для $|\rho(x, x_0)|$ подставили квадрат расстояния $(\|x - x_0\|^2)$ (такой переход от расстояния к его квадрату хорошо бы обосновать монотонностью оптимизируемой функции, но мы это опустим), а также перед коэффициентом λ для удобства поставили 2, что несущественно.

Далее надо приравнять нуль частные производные лагранжиана по его аргументам. Из $\frac{\partial \mathcal{L}}{\partial x} = 0$ получаем: $2(x - x_0) + 2\lambda w = 0 \Rightarrow x = x_0 - \lambda w$.

Теперь домножим это уравнение скалярно на $\langle w \rangle$ и выразим λ :

$$\langle w, x \rangle = \langle w, x_0 \rangle - \lambda \langle w, w \rangle \Rightarrow \lambda = \frac{\langle w, x_0 \rangle - \langle w, x \rangle}{\|w\|^2}$$

Тогда наконец

$$\|x - x_0\| = \|(x_0 - \lambda w) - x_0\| = \|\lambda w\| = |\lambda| \|w\| = \frac{|\langle w, x_0 \rangle - \langle w, x \rangle|}{\|w\|}$$

Таким образом, линейный классификатор разделяет пространство на две части с помощью гиперплоскости, и при этом одно полупространство относит к положительному классу, а другое – к отрицательному.

Пожалуй, самый известный и популярный на практике представитель семейства линейных классификаторов – логистическая регрессия. На русском языке про нее можно почитать в [статье](#) открытого курса по машинному обучению или в упомянутых [лекциях](#) Евгения Соколова. В этих лекциях также объясняется, как происходит обучение модели (подбор весов $\|w\|$) за счет минимизации функции потерь.

Оказывается полезным рассмотреть выражение $\langle M(x_i, y_i), w \rangle = y_i \langle w, x_i \rangle$.

Это **отступ классификации** (margin) для объекта обучающей выборки (x_i, y_i) . К сожалению, его очень легко перепутать с зазором классификации, который появится чуть ниже в изложении интуиции метода опорных векторов. Чтобы не путать: отступ определен на конкретном объекте обучающей выборки.

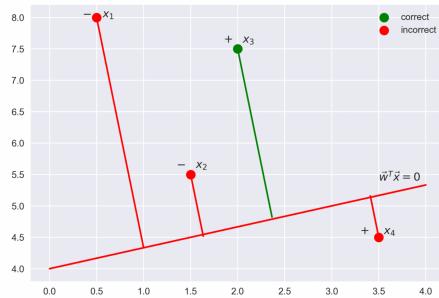


Fig. 104 Иллюстрация к понятию отступа классификации

Отступ – это своего рода “уверенность” модели в классификации объекта (x_i, y_i) :

- если отступ большой (по модулю) и положительный, это значит, что метка класса поставлена правильно, а объект находится далеко от разделяющей гиперплоскости (такой объект классифицируется уверенно). На рисунке – (x_3) .
- если отступ большой (по модулю) и отрицательный, значит метка класса поставлена неправильно, а объект находится далеко от разделяющей гиперплоскости (скорее всего такой объект – аномалия, например, его метка в обучающей выборке поставлена неправильно). На рисунке – (x_1) .
- если отступ малый (по модулю), то объект находится близко к разделяющей гиперплоскости, а знак отступа определяет, правильно ли объект классифицирован. На рисунке – (x_2) и (x_4) .

Далее увидим, что понятие отступа классификации – часть функции потерь, которая оптимизируется в методе опорных векторов.

Интуиция метода опорных векторов

Метод опорных векторов (Support Vector Machine, SVM) основан на идее максимизации зазора между классами. Пока не вводим этот термин формально, но передадим интуицию метода. На Рис. [Fig. 105](#) показана линейно-разделимая выборка, кружки соответствуют положительным примерам, а квадраты – отрицательным (или наоборот), а оси – некоторым признакам этих примеров. На рисунке слева показаны две прямые (в общем случае – гиперплоскости), разделяющие выборку. Кажется, что синяя прямая лучше тем, что она дальше отстоит от примеров обучающей выборки, чем красная прямая (зазор – больше), и потому лучше будет разделять другие примеры из того же распределения, что и примеры обучающей выборки. То есть такой линейный классификатор будет лучше обобщаться на новые данные. Теория подтверждает описанную интуицию [[MRT18](#)].

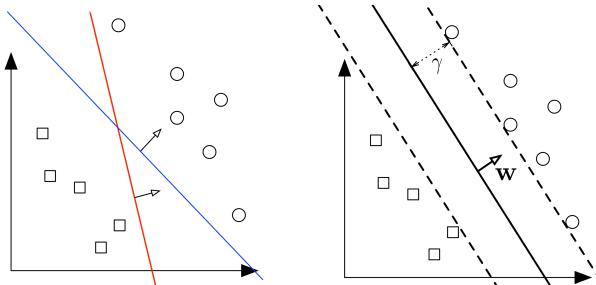


Fig. 105 Слева показаны две прямые (в общем случае – гиперплоскости), разделяющие выборку.

Справа показана прямая, максимизирующая зазор между классами. Источник: [лекция](#) Cornell

 Note

Одним из ключевых авторов алгоритма SVM является Владимир Вапник – советский и американский (с 1991-го года) ученый, который также сделал огромный вклад в теорию классического машинного обучения. Его имя носит одно из ключевых теоретических понятий машинного обучения – размерность Вапнича-Червоненкиса.

Метод опорных векторов для линейно-разделимой выборки

Будем рассматривать линейные классификаторы вида

$\llbracket a(x) = \text{sign}(\langle w, x \rangle + b), \qquad w \in \mathbb{R}^d, b \in \mathbb{R}. \rrbracket$

Заметьте, что мы вернули сдвиг (*bias*) $\backslash(b)$. Будем считать, что существуют такие параметры $\backslash(w_*)$ и $\backslash(b_*)$, что соответствующий им классификатор $\backslash(a(x))$ не допускает ни одной ошибки на обучающей выборке. В этом случае говорят, что выборка **линейно разделима**.

Заметим, что если одновременно умножить параметры $\backslash(w)$ и $\backslash(b)$ на одну и ту же положительную константу, то классификатор не изменится. Распорядимся этой свободой выбора и отнормируем параметры так, что

$\| \min_{x \in X} \| \langle w, x \rangle + b \| = 1. \]$

Как мы увидели выше, расстояние от произвольной точки $(x_0 \in \mathbb{R}^d)$ до гиперплоскости, определяемой данным классификатором, равно

$$\rho(x_0, a) = \frac{\langle w, x \rangle + b}{\|w\|}.$$

Тогда расстояние от гиперплоскости до ближайшего примера из обучающей выборки равно

$$\forall \{ \min_{x \in X} \frac{\|w^T x + b\|}{\|w\|} = \frac{1}{\|w\|} \min_{x \in X} \|w^T x + b\| = \frac{1}{\|w\|} \|w\|.$$

Данная величина также называется **зазором** (margin). Опять же, эту величину очень легко перепутать с отступом классификации, про который мы говорили выше и который тоже margin в англоязычном варианте. Заметим, что отступ – функция параметров $\|w\|$ и конкретного примера обучающей выборки $\|(x_i, y_i)\|$, а зазор – функция только параметров $\|w\|$ при описанном трюке с нормировкой $\|w\|$ и $\|b\|$ (в противном случае зазор – также функция сдвига $\|b\|$ и всех примеров $\|x\|$).

Таким образом, если классификатор без ошибок разделяет обучающую выборку, то ширина его разделяющей полосы равна $\|\frac{1}{\sqrt{2}}(\mathbf{w}^T \mathbf{x})\|$. Максимизация ширины разделяющей полосы приводит к повышению обобщающей способности классификатора [MRT18]. На повышение обобщающей способности направлена и регуляризация, которая штрафует большую норму весов – а чем больше норма весов, тем меньше ширина разделяющей полосы.

Итак, требуется построить классификатор, идеально разделяющий обучающую выборку, и при этом имеющий максимальный отступ.

Запишем соответствующую оптимизационную задачу, которая и будет определять метод опорных векторов для линейно разделимой выборки (hard margin support vector machine):

```
\begin{split} \begin{equation} \label{eq:svmSep} \left\langle \begin{aligned} & \frac{1}{2} \|w\|^2 + \min_{\{w, b\}} \\ & \langle w, x_i \rangle + b \end{aligned} \right\rangle \geq 1, \quad i = 1, \dots, |E|. \end{aligned} \right. \end{equation} \end{split}
```

Здесь мы воспользовались тем, что линейный классификатор дает правильный ответ на примере $\langle x_i \rangle$ тогда и только тогда, когда $\langle y_i \rangle (\langle w, x_i \rangle + b) \geq 0$ (вспомним, что $\langle M_i = y_i (\langle w, x_i \rangle + b) \rangle$ – отступ классификации для примера $\langle (x_i, y_i) \rangle$ обучающей выборки). Более того, из условия нормировки $\langle \min_{x \in X} \langle w, x \rangle + b \rangle = 1$ следует, что $\langle y_i (\langle w, x_i \rangle + b) \rangle \geq 1$.

В данной задаче функционал является строго выпуклым, а ограничения линейными, поэтому сама задача является выпуклой и имеет единственное решение. Более того, задача является квадратичной и может быть решена крайне эффективно.

Метод опорных векторов для линейно-неразделимой выборки

Рассмотрим теперь общий случай, когда выборку невозможно идеально разделить гиперплоскостью. Это означает, что какие бы $\{w\}$ и $\{b\}$ мы ни взяли, хотя бы одно из ограничений в предыдущей задаче будет нарушено:

$\forall \exists x \in X: \forall y \in \text{left}(\langle w, x \rangle \rangle + b \leq 1)$

Сделаем эти ограничения **мягкими**, введя штраф $\lambda(x_i \neq \text{def } 0)$ за их нарушение:

$$\forall y \ i \ \leftarrow (\langle w, x \rangle \ i \ \rangle + b) \geq 1 - \epsilon \quad i = 1, \dots, n.$$

Отметим, что если отступ объекта лежит между нулем и единицей ($0 \leq y_i \leftarrow \langle w, x_i \rangle + b < 1$), то объект верно классифицируется, но имеет ненулевой штраф ($x_i > 0$). Таким образом, мы штрафуем объекты за попадание внутрь разделяющей полосы.

Величина $\frac{1}{2} \|w\|^2$ в данном случае называется **мягким зазором** (soft margin). С одной стороны, мы хотим максимизировать зазор, с другой – минимизировать штраф за неидеальное разделение выборки $\sum_i \ell(y_i - f(x_i))$.

Эти две задачи противоречат друг другу: как правило, излишняя подгонка под выборку приводит к маленькому зазору, и наоборот – максимизация зазора приводит к большой ошибке на обучении. В качестве компромисса будем минимизировать взвешенную сумму двух указанных величин.

Приходим к оптимационной задаче, соответствующей методу опорных векторов для линейно неразделимой выборки (soft margin support vector machine):

$$\begin{aligned} & \text{min}_{w, b} \|w\|^2 + C \sum_i \ell(y_i - \langle w, x_i \rangle) \\ & \text{subject to } y_i(\langle w, x_i \rangle + b) \geq 1 \quad \forall i \end{aligned}$$

Чем больше здесь параметр C , тем сильнее мы будем настраиваться на обучающую выборку. Данная задача также является выпуклой и имеет единственное решение.

Сведение к безусловной задаче оптимизации

Покажем, что задачу метода опорных векторов можно свести к задаче безусловной оптимизации функционала, который имеет вид верхней оценки на долю неправильных ответов.

Перепишем условия задачи:

$$\begin{aligned} & \text{min}_{w, b} \|w\|^2 + C \sum_i \max(0, 1 - y_i(\langle w, x_i \rangle + b)) \\ & \text{subject to } y_i(\langle w, x_i \rangle + b) \geq 1 \quad \forall i \end{aligned}$$

Поскольку при этом в функционале требуется, чтобы штрафы $\ell(y_i)$ были как можно меньше, то можно получить следующую явную формулу для них:

$$\ell(y_i) = \max(0, 1 - y_i(\langle w, x_i \rangle + b))$$

Данное выражение для $\ell(y_i)$ уже учитывает в себе все ограничения задачи, описанной выше. Значит, если подставить его в функционал, получим безусловную задачу оптимизации:

$$\text{min}_w \|w\|^2 + C \sum_i \max(0, 1 - y_i(\langle w, x_i \rangle + b))$$

Эта задача является негладкой, поэтому решать её может быть достаточно тяжело. Тем не менее, она показывает, что метод опорных векторов, по сути, как и логистическая регрессия, строит верхнюю оценку на долю ошибок и добавляет к ней стандартную квадратичную регуляризацию. Только если в случае логистической регрессии этой верхней оценкой была логистическая функция потерь (опять сделаем отсылку к [статье](#) из открытого курса машинного обучения), то в случае метода опорных векторов это функция вида $L(y, z) = \max(0, 1 - yz)$, которая называется **кусочно-линейной функцией потерь (hinge loss)**.

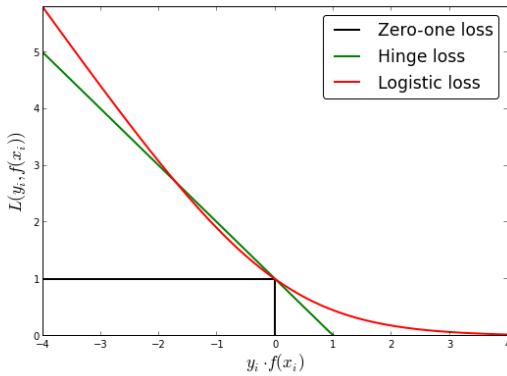


Fig. 106 Пороговая, кусочно-линейная и логистическая функции потерь

Это становится понятнее в терминах упомянутого выше отступа $(M_i = y_i(\langle w, x_i \rangle + b))$ на примере обучающей выборки. В идеале мы хотели бы штрафовать классификатор за ошибку на примере: $(L_{1/0}(M_i) = [M_i < 0])$. Это пороговая функция потерь (zero-one loss), ее график изображен черным на Fig. 106 как функция от отступа. К сожалению, напрямую мы не можем эффективно оптимизировать такую функцию градиентными методами из-за разрыва в нуле, поэтому оптимизируется верхняя оценка zero-one loss. В случае логистической регрессии – логистическая функция потерь $L(M_i) = \log(1 + e^{-M_i})$ (красная на рисунке выше), а в случае метода опорных векторов – кусочно-линейная функция $L(M_i) = \max(0, 1 - M_i)$ (зеленая на рисунке выше).

Note

Бытует мнение, что метод опорных векторов сегодня нигде не используется из-за его сложности (как минимум квадратичной по числу примеров). Однако, это не так. Линейный SVM вполне неплохо можно применять в задачах с высокой размерностью объектов обучающей выборки, например, для классификации текстов с [Tf-IDF](#) или любым другим разреженным представлением. В частности, [Vowpal Wabbit](#) – очень эффективная утилита для решения многих задач машинного обучения – по умолчанию использует hinge-loss для задач классификации, то есть по сути в этом сценарии применения является линейным SVM. Кусочно-линейная функция потерпеть хорошо тем, что у нее очень простая производная – положительная константа либо ноль. Это удобно использовать с SGD и большими выборками, когда приходится делать миллиарды обновлений весов.

Про прелести Vowpal Wabbit и обучение на гигабайтах данных за считанные минуты можно почитать в [статье](#) открытого курса машинного обучения.

Решение задачи метода опорных векторов

Итак, метод опорных векторов сводится к решению задачи оптимизации

$$\begin{aligned} \min_w \left\| w \right\|^2 + C \sum_i \max(0, 1 - y_i(w^T x_i + b)) \end{aligned}$$

Для решения таких *условных* задач оптимизации с условиями в виде неравенств или равенств часто используют лагранжиан и двойственную задачу оптимизации. Этот подход исчерпывающе описан в классической книге Бойда по оптимизации [[BV04](#)], а на русском языке можно обратиться к [конспекту](#) Евгения Соколова. Также для понимания материала рекомендуется рассмотреть [“игрушечный” пример](#) решения задачи метода опорных векторов в случае линейно-разделимой выборки из 5 примеров.

Построим двойственную задачу к задаче метода опорных векторов. Запишем лагранжиан:

$$L(w, b, \lambda) = \frac{1}{2} \|w\|^2 + C \sum_i \max(0, 1 - y_i(w^T x_i + b))$$

Выпишем условия Куна-Таккера:

$$\nabla_w L = w - \sum_i y_i x_i = 0 \quad (11)$$

$$\nabla_b L = \sum_i y_i = 0 \quad (12)$$

$$\lambda_i \geq 0, \lambda_i y_i = 0 \quad (13)$$

$$w^T x_i + b = y_i \quad (14)$$

$$y_i(\lambda_i w^T x_i + b) = 0 \quad (15)$$

Проанализируем полученные условия. Из (11) следует, что вектор весов, полученный в результате настройки SVM, можно записать как линейную комбинацию объектов, причем веса в этой линейной комбинации можно найти как решение двойственной задачи.

В зависимости от значений λ_i и μ_i объекты x_i разбиваются на три категории:

- $\lambda_i = 0$, $\mu_i = 0$. Такие объекты не влияют решение w (входят в него с нулевым весом λ_i), правильно классифицируются ($y_i \lambda_i = 0$) и лежат вне разделяющей полосы. Объекты этой категории называются *периферийными*.
- $\lambda_i > 0$, $\mu_i < C$. Из условия (14) следует, что $y_i \lambda_i = 1$, то есть объект лежит строго на границе разделяющей полосы. Поскольку $\lambda_i > 0$, объект влияет на решение w . Объекты этой категории называются *опорными граничными*.
- $\lambda_i > 0$, $\mu_i = C$. Такие объекты могут лежать внутри разделяющей полосы ($0 < \lambda_i < 2$) или выходить за ее пределы ($\lambda_i \geq 2$). При этом если $\lambda_i < 1$, то объект классифицируется правильно, в противном случае – неправильно. Объекты этой категории называются *опорными нарушителями*.

Отметим, что варианта $\lambda_i > 0$, $\mu_i < C$ быть не может, поскольку при $\lambda_i > 0$ из условия дополняющей нежесткости (15) следует, что $\mu_i = 0$, и отсюда из уравнения (13) получаем, что $\lambda_i = C$.

Итак, итоговый классификатор зависит только от объектов, лежащих на границе разделяющей полосы, и от объектов-нарушителей (с $\lambda_i > 0$).

Построим двойственную функцию. Для этого подставим выражение (11) в лагранжиан и воспользуемся уравнениями (12) и (13) (даные три уравнения выполнены для точки минимума лагранжиана при любых фиксированных λ и μ):

$$\begin{aligned} \begin{aligned} L &= \frac{1}{2} \left(\sum_{i=1}^n \lambda_i y_i x_i - \sum_{i,j=1}^n \lambda_i y_i x_i \right)^2 - \sum_{i,j=1}^n \lambda_i y_i x_i \\ &\quad - \lambda_i \sum_j \lambda_j y_j x_j \langle x_i, x_j \rangle - b \underbrace{\left(\sum_{i=1}^n \lambda_i y_i \right)_0}_0 + \sum_{i=1}^n \lambda_i y_i \sum_{j=1}^n \lambda_j y_j \langle x_i, x_j \rangle \end{aligned} \end{aligned}$$

Мы должны потребовать выполнения условий (12) и (13) (если они не выполнены, то двойственная функция обращается в минус бесконечность), а также неотрицательность двойственных переменных ($\lambda_i \geq 0$), ($\mu_i \geq 0$). Ограничение на (λ_i) и условие (13), можно объединить, получив ($\lambda_i \leq C$). Приходим к следующей двойственной задаче:

$$\begin{aligned} \begin{aligned} \text{Left} \{ \begin{aligned} &\sum_{i=1}^n \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle - b \max_{\lambda} \lambda_i \leq 0 \quad \forall i = 1, \dots, n \quad \& \sum_{i=1}^n \lambda_i = 1 \end{aligned} \end{aligned}$$

Она также является вогнутой, квадратичной и имеет единственный максимум.

Ядерный переход

Двойственная задача SVM (17) зависит только от скалярных произведений объектов – отдельные признаковые описания никак не входят в неё.

Note

Обратите внимание, как много это значит: решение SVM зависит только от скалярных произведений объектов (то есть *похожести*, если упрощать), но не от их признаковых описаний объектов. Это значит, что метод обобщается и на те случаи, когда признаковых описаний объектов нет или их получить очень дорого, но зато есть способ задать расстояние (то есть “измерить сходство”) между объектами.

Значит, можно сделать ядерный переход:

$$\begin{aligned} \begin{aligned} \text{Left} \{ \begin{aligned} &\sum_{i=1}^n \lambda_i \lambda_j y_i y_j K(x_i, x_j) - b \max_{\lambda} \lambda_i \leq 0 \quad \forall i = 1, \dots, n \quad \& \sum_{i=1}^n \lambda_i = 1 \end{aligned} \end{aligned}$$

Здесь ($K(x_i, x_j)$) – это функция-ядро, определенная на парах векторов, которая должна быть симметричной и неотрицательно определенной ([теорема Мерсера](#)).

Вернемся к тому, какое представление классификатора дает двойственная задача. Из уравнения (11) следует, что вектор весов (w) можно представить как линейную комбинацию объектов из обучающей выборки. Подставляя это представление (w) в классификатор, получаем

$$a(x) = \text{sign} \left(\sum_{i=1}^n \lambda_i y_i \langle x_i, x \rangle + b \right)$$

Таким образом, классификатор измеряет сходство нового объекта с объектами из обучающей выборки, вычисляя скалярное произведение между ними. Это выражение также зависит только от скалярных произведений, поэтому в нём тоже можно перейти к ядру.

Note

Опять подчеркнем, что классификация нового примера зависит только от скалярных произведений – “похожести” нового примера на примеры из обучающей выборки, и то не все, а только опорные.

Note

В указанном выше представлении фигурирует переменная сдвига (b), которая не находится непосредственно в двойственной задаче. Однако ее легко восстановить по любому граничному опорному объекту (x_i), для которого выполнено ($y_i = 1, 0 < \lambda_i < C$). Для него выполнено ($y_i \langle w, x_i \rangle + b = 1$), откуда получаем

$$b = y_i - \langle w, x_i \rangle$$

Как правило, для численной устойчивости берут медиану данной величины по всем граничным опорным объектам:

$$b = \text{med } y_i - \langle w, x_i \rangle, \quad y_i = 1, 0 < \lambda_i < C$$

[Fig. 107](#) – пожалуй, самый известный рисунок в контексте SVM, он иллюстрирует ядерный трюк, в свою очередь, одну из самых красивых идей в истории машинного обучения. За счет ядерного перехода можно достигнуть линейной разделимости выборки даже в том случае, когда исходная обучающая выборка не является линейно разделимой.

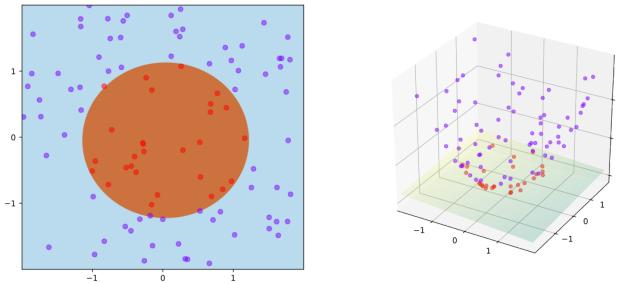


Fig. 107 Пример разделимости в новом пространстве

Наиболее часто используемые ядра:

- Линейное $\langle K(x, y) = \langle x, y \rangle \rangle$ – по сути, линейный SVM, рассмотренный выше;
- Полиномиальное ядро $\langle K(x, y) = (\langle x, y \rangle + c)^d \rangle$, определенное для степени ядра d и параметра нормализации c ;
- Гауссово ядро, также известное как RBF (radial-basis functions) $\langle K(x, y) = e^{-\frac{\|x - y\|^2}{\sigma^2}} \rangle$ с параметром ядра σ .

Плюсы и минусы SVM

Плюсы:

- хорошо изучены, есть важные теоретические результаты;
- красиво формулируется как задача оптимизации;
- линейный SVM быстрый, может работать на очень больших выборках;
- линейный SVM так же хорошо интерпретируется, как и прочие линейные модели;
- решение зависит только от скалярных произведений векторов, а идея “ядерного трюка” – одно из самых красивых в истории машинного обучения;
- нелинейный SVM обобщается на работу с самыми разными типами данных (последовательности, графы и т.д.) за счет специфичных ядер.

Минусы:

- нелинейный SVM имеет высокую вычислительную сложность и принципиально плохо масштабируется (оптимизационную задачу нельзя “решить на подвыборках” и как-то объединить решения);
- нелинейный SVM по сути не интерпретируется (“black box”);
- в задачах классификации часто хочется выдать вероятность отнесения к классу, SVM это не умеет делать, а эвристики, как правило, приводят к плохо откалиброванным вероятностям;
- ядерный SVM уступает специфичным нейронным сетям уже во многих задачах, например, в приложениях к графикам

О блоке “Квантовые нейронные сети”

Этот блок включает в себя обзор:

- квантовых нейронных сетей;
- сверточных вариационных квантовых схем.

Продвинутая тема блока дополнительно рассказывает о плато Баррена.

Квантовые нейронные сети

Описание лекции

В этой лекции мы пройдёмся по расширению идеи нейронных сетей на квантовые компьютеры – мы уже прошли и [вариационные квантовые схемы \(VQC\)](#), и комбинацию [квантовых и классических градиентов в них](#) в соответствующих блоках. Всё что осталось – это объединить всё изученное в общую картину и заняться

обучением этих самых квантовых нейронных сетей. Для того, чтобы лучше разобраться в том, какие ограничения и возможности предоставляет этот подход, мы приведём несколько публикаций о квантовых и гибридных генеративных состязательных сетях.

Введение

Как уже было упомянуто в лекции по VQC, на данный момент квантовые вычислители ещё недостаточно развиты для того, чтобы в одиночку решать большие задачи, имеющие практическое значение для индустрии – это в особенной степени актуально для нейронных сетей, которые и в классическом сценарии требуют значительных вычислительных ресурсов. Именно поэтому на данный момент наиболее популярна категория гибридных вариационных алгоритмов, которые обучают квантовую параметрическую схему (**QNN**) при помощи классической оптимизации, например, **VQ Eigensolvers** и **Quantum Approximate Optimization Algorithms**. В общем и целом идея гибридных алгоритмов заключается в оптимизации над некоторым классом параметрических вычислений для минимизации энергии волновой функции (**VQE/QAOA**), экстракции нелокальной информации (**QNN Classifiers**) или генерации данных, соответствующих квантовому распределению (**Quantum Generative Models**).

Применение

Лучше всего понять адекватность и применимость какой-то технологии, особенно основанной на комплексном научном базисе, позволяет как раз таки ее применение - чего удалось добиться кому-то на практике и насколько результаты пригодны для того, чтобы хвататься за технологию. Вместе с тем, практические результаты дадут нам глубже понять и в последствии объяснить как работают гибридные сети под капотом. Приведём два примера, которые хорошо показывают разницу в ограничениях при использовании полностью квантовых и гибридных квантово-классических сетей.

В первом примере была построена полностью квантовая генеративная состязательная сеть, задачей которой являлось воспроизвести MNIST, однако у всего эксперимента был один нюанс. Из-за текущего размера квантовых вычислителей было предложено уменьшить размерность MNIST с 784 до 4 при помощи РСА, чтобы его хоть как-то можно было отправить в квантовую схему. Очевидно, сеть такого размера не в состоянии тягаться с классическими сетями на десятки тысяч весов, поэтому конечным результатом стало сравнение по количеству весов при идентичных результатах. Полностью квантовая сеть смогла получить такие же результаты, как и классическая, при этом имея на 95% меньше параметров. Эти результаты всё ещё поднимают вопрос того, будут ли они справедливыми для моделей, адекватно справляющихся с задачей.

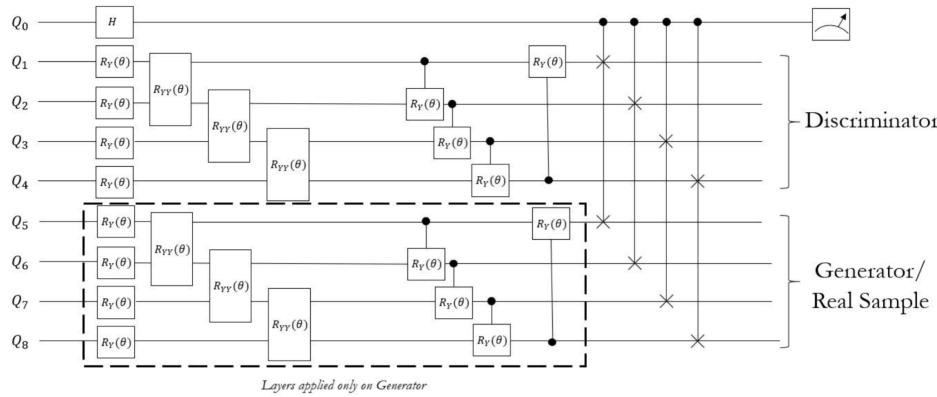


Fig. 108 Схема квантовой генеративной состязательной сети из публикации.

Источник публикации

Во втором примере была построена гибридная квантово-классическая генеративная состязательная сеть, задачей которой являлось воспроизвести QM-9, популярный датасет молекул до 9 атомов, представленный молекулярным графом в виде матрицы 9 на 9 и атомарным вектором из 9 элементов. Генеративная часть модели состояла из нескольких квантовых схем и, по большей части, классической сети, предсказывающей узлы и связи между ними. Дискриминативная часть была полностью классической во всех вариациях модели. В конечном итоге им удалось добиться снижения числа параметров по сравнению с полностью классической сетью на 85% и, при увеличении числа кубитов, на 98%. Таким образом они показали, что комбинация классических и квантовых схем должна соответствовать той же самой закономерности, что и полностью квантовые сети.

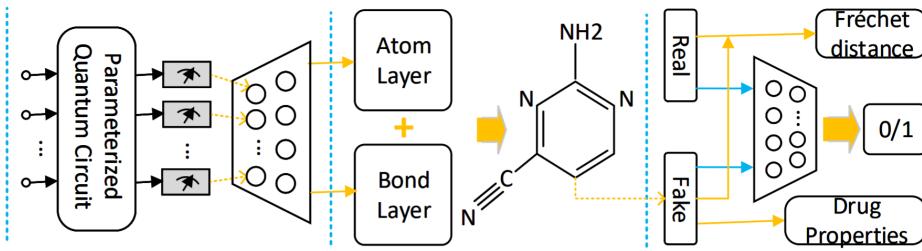


Fig. 109 Схема гибридной генеративной состязательной сети из публикации.

Источник публикации

За счёт большей актуальности на текущих практических задачах именно последнего, гибридного подхода, дальнейшая часть лекции будет рассматривать именно его.

Архитектура

В идеале этот подход подразумевал бы, что при помощи классического оптимизатора мы обучаем некоторую параметрическую схему на квантовом вычислителе, однако в текущих реалиях NISQ этот подход является невозможным, поэтому большая часть параметрической схемы остаётся на классических вычислителях. В данном блоке мы поговорим о подходе, связанном с **QNN Classifiers**, которые следуют вышеупомянутому принципу и обучаются градиентным спуском практически так же, как и обычные классические сети, позволяя градиенту протекать между квантовой и классической частью сети.

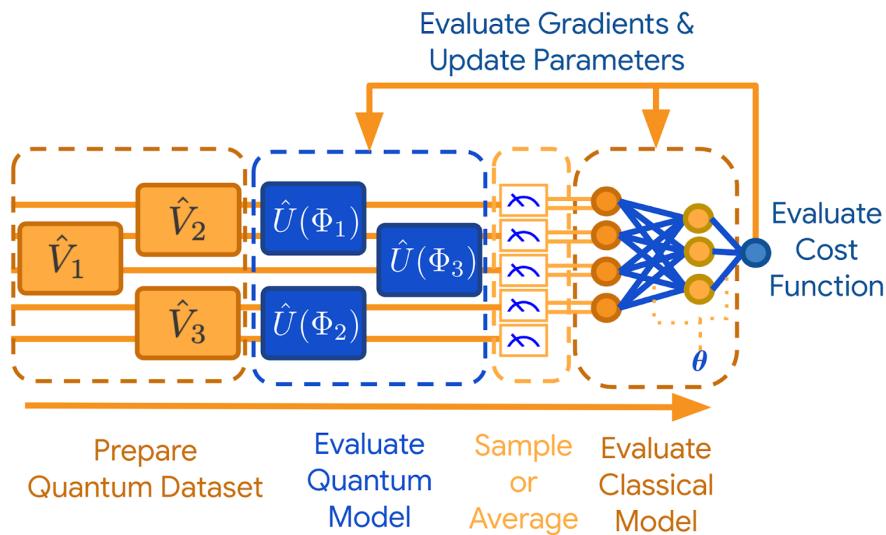


Fig. 110 Схема обучения гибридной нейронной сети из [BVM+21]

На изображении гибридной сети процедура практически идентична классическому обучению сетей, в котором добавляется процесс кодирования классических данных в квантовые операторы, и процесс измерения квантового состояния для того, чтобы передать уже классическую информацию для дальнейших вычислений на классическом устройстве, как это было описано в лекции по **VQC**.

Анзац

Зачастую в литературе по **VQC**, особенно когда речь идёт о нейронных сетях, упоминается такая вещь как **ansatz** – по своей сути это заранее подготовленные участки параметрической схемы, которые могут быть использованы как составные блоки сети. Если проводить параллели с классическим машинным обучением, то в рамках библиотеки **PennyLane** эти схемы называются **templates** и могут представлять собой, например, свёрточный слой или эмбеддинг, но и более общие элементы квантовой схемы вроде подготовки состояний или перестановок между кубитами. Более подробно мы остановимся на них слегка позже в курсе, в одной из следующих глав, а пока что в общих чертах пройдёмся по обучению квантовых нейронных сетей без сложностей внутренней кухни.

Note

Интересно, но термин *ansatz* пришел в квантовые вычисления и QML из теоретической физики. Этот термин имеет немецкое происхождение, так как в первой половине XX века именно немецкие научные журналы были самыми топовыми. Частое употребление этого термина в отношении квантового машинного обучения объясняется тем, что большая часть специалистов в этой области это именно люди, занимающиеся теоретической физикой.

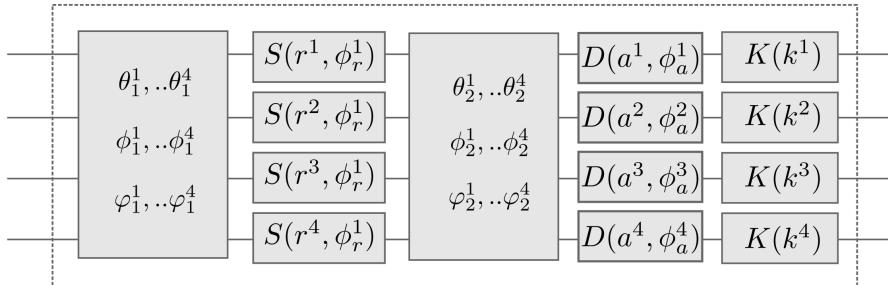


Fig. 111 Ansatz, соответствующий свёрточному слою нейронной сети в PennyLane.

[Источник картинки](#)

ФУНКЦИЯ ПОТЕРЬ

Функция потерь работает таким же образом, как и в полностью классических сетях, так как оптимизация происходит на классическом железе – единственное, что отличается, – это объединение квантовых и классических градиентов. Градиент по нашей квантовой схеме получается при помощи замера состояния, которое может варьироваться из-за вероятностной природы кубита, поэтому несколько замеров позволяют аппроксимировать ожидаемый градиент при помощи методов вроде [finite differences](#) или [parameter-shift](#), после чего остаётся только совместить его с классическим.

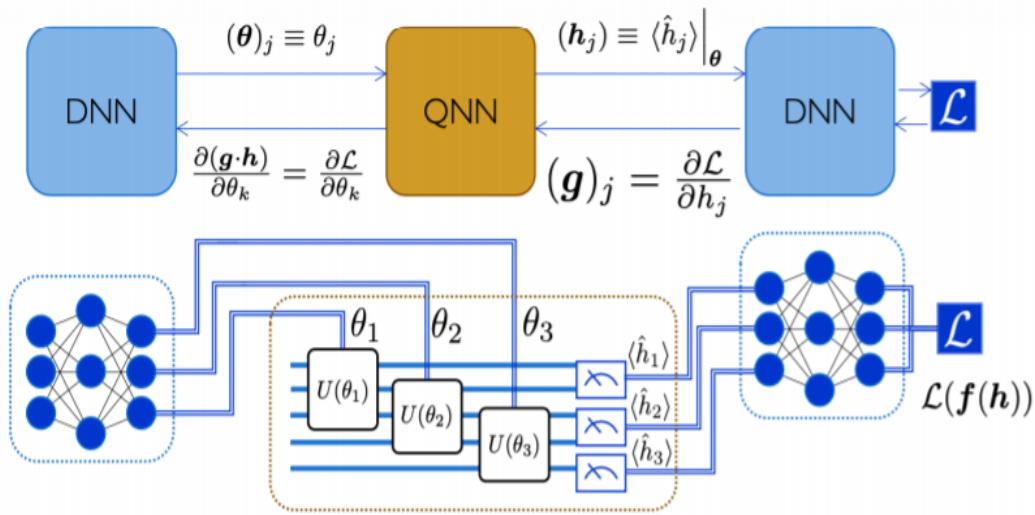


Fig. 112 Распространение градиентов от функции потерь в гибридной схеме.

Сеть от начала до конца

В конечном итоге мы имеем следующую последовательность действий для того, чтобы собрать гибридную нейронную сеть:

- трансформировать данные из классических в квантовые представления;
- отправить эти данные для вычисления на квантовой схеме;

- просэмплировать и замерить результат квантовой схемы;
- отправить результаты для вычисления на классической схеме;
- оценить ошибку, рассчитать градиенты и обновить параметры.

Что мы узнали из лекции

- В ближайшие годы полностью квантовые нейронные сети не смогут решать задачи целиком, поэтому будут использоваться в качестве составляющей гибридного квантово-классического решения.
- Так же как и для полностью квантовых сетей, гибридные сети позволяют уменьшить количество необходимых параметров по сравнению с полностью классическими сетями.
- Обучение подобных сетей практически идентично обучению классических сетей за исключением нескольких трюков, необходимых для работы с параметрами квантовых схем.

Сверточные вариационные квантовые схемы

Описание лекции

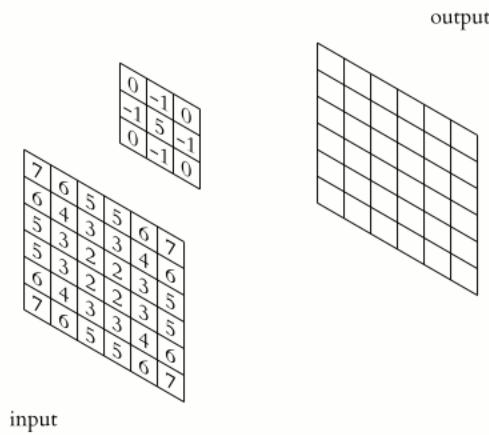
На этой лекции мы рассмотрим аналог сверточных слоев нейронной сети и приведем пример их использования.

Лекция расскажет:

- о том как работают “классические” свертки в нейронных сетях;
- что такое сверточные вариационные квантовые схемы;
- как их можно использовать для задач машинного обучения на примере датасета MNIST.

Сверточные нейронные сети

Данный тип нейронных сетей нацелен на обработку изображений и представляет из себя последовательный набор **сверточных слоев**, чередующиеся с другими вспомогательными слоями, например, с функциями нелинейного преобразования (активациями). Пример работы **сверточного слоя** представлен на рисунке ниже.



Свое название **сверточный слой** получил из-за наличия операции свертки (конволюции), суть которой в том, что каждый фрагмент входного изображения (input) умножается на матрицу (ядро размера 3x3, kernel) свертки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения (output).

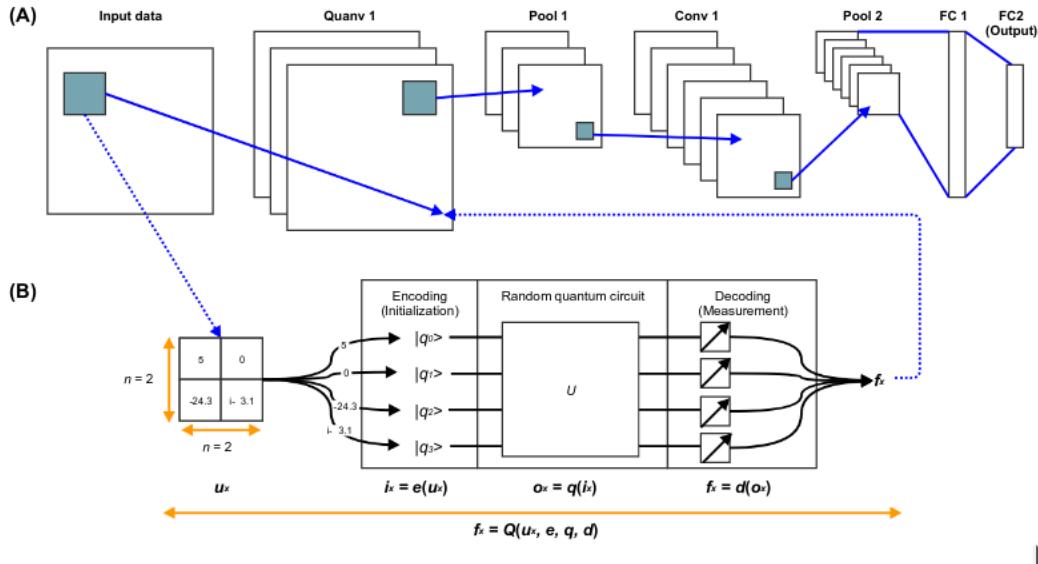
Сверточные слои являются ключевыми элементами алгоритмов по распознаванию и классификации изображений. Большинство успехов в области компьютерного зрения за последнее десятилетие связано именно с ними.

Квантовая свертка

Активное развитие квантовых вычислений создает запрос на построение квантового аналога свертки. Работа ([Henderson M. et.al., 2019](#)), описывающая подобный подход, предлагает следующую схему реализации на основе вариационных квантовых схем:

- изображение разбивается на блоки размера 2x2;

- для каждого блока осуществляется кодирование входного сигнала к квантовому состоянию, которое в данной лекции осуществляется параметризованным вращением кубита из основного состояния;
- квантовые вычисления выполняются над входными кубитами посредством случайной квантовой цепи;
- проводится измерение квантовой системы для получения классических величин;
- выполняя данную процедуру для всех 2×2 областей на выходе получается многоканальное изображение (трехканальное на примере ниже), которое будет использоваться в следующих слоях классической или квантовой нейронной сети.



Основное отличие по отношению к классической свертке состоит в том, что квантовая схема может генерировать очень сложные ядра, вычисления которых могут быть классически невоспроизводимыми. Указанные сложные ядра, а именно ядра в высокоразмерном пространстве Гильберта, могут дать преимущества квантовым сверткам по отношению к классическим.

Пример реализации и работы такого блока квантовой свертки будет продемонстрирован ниже.

Подготовка модели и данных

Пример кода реализации гибридной квантово-классической сети с квантовым сверточным слоем взят из [документации](#) библиотеки `PennyLane` с небольшими изменениями.

Сначала загружаем библиотеки `PennyLane` для квантовых вычислений, `tensorflow` для обучения классификатора и `matplotlib` для отрисовки результатов.

```
import pennylane as qml
from pennylane import numpy as np
from pennylane.templates import RandomLayers

import os
os.environ['TF_CPP_MIN_LOG_LEVEL'] = '2'
import tensorflow as tf
from tensorflow import keras

import matplotlib.pyplot as plt
```

Устанавливаем параметры модели для обучения: число эпох, число слоев квантовой свертки и его выходных каналов, а также число тренировочных и тестовых примеров. Число примеров выбрано малым для увеличения скорости демонстрации. Также фиксируем инициализацию генераторов случайных чисел для `numpy` и `tensorflow`.

```
n_epochs = 20    # Количество эпох обучения
n_layers = 1     # Число случайных квантовых блоков
n_wires = 3      # Число выходных каналов после квантовых блоков
n_train = 20      # Размер тренировочного датасета
n_test = 10       # Размер тестового датасета

# Инициализация генераторов случайных чисел
np.random.seed(0)
tf.random.set_seed(255)
```

Для демонстрации используем датасет MNIST, который создан для предсказания цифры (от 0 до 9) по его изображению

```
mnist_dataset = keras.datasets.mnist
(train_images, train_labels), (test_images, test_labels) = mnist_dataset.load_data()

# Ограничение размера датасета
train_images = train_images[:n_train]
train_labels = train_labels[:n_train]
test_images = test_images[:n_test]
test_labels = test_labels[:n_test]

# Нормализация изображений из диапазона (0, 255) в (0, 1)
train_images = train_images / 255
test_images = test_images / 255

# Добавление дополнительной размерности к данным для сверточных каналов
train_images = np.array(train_images[...], tf.newaxis), requires_grad=False)
test_images = np.array(test_images[...], tf.newaxis), requires_grad=False)
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/mnist.npz
```

```
8192/11490434 [.....] - ETA: 0s
```

```
8806400/11490434 [=====>.....] - ETA: 0s
```

```
11493376/11490434 [=====] - 0s 0us/step
```

Случайная квантовая цепь

В качестве блока квантовой свертки используется **случайная квантовая цепь** (RandomLayers в PennyLane). Блок реализуется путем случайного выбора для части кубитов операций вращения, а для части пар кубитов парных квантовых гейтов. На последнем этапе генерации блока осуществляется случайное перемешивание очередности применения операций.

В нашем примере случайная квантовая цепь обрабатывает 4 кубита. Таким образом часть из 4 кубитов получит какие-либо операции вращения, а часть из 10 пар кубитов - гейты. Финальная очередьность операций будет случайной.

Следует заметить, что в нашем примере используется фиксированный (необучаемый) блок случайной квантовой цепи. Он будет использован как блок препроцессинга. В будущем, когда в библиотеке будет реализована возможность расчета градиентов и изменения параметров блока на их основании, указанный блок тоже можно будет обучать.

Реализация квантовой свертки

Далее создаем устройство, симулирующее работу 3 кубитов. Оно состоит из трех частей:

- кодировщик, преобразующий входные данные в квантовые состояния с помощью RY поворотов;
- случайная квантовая цепь;
- измерения, дающие 3 выходных значения.

```
dev = qml.device("default.qubit", wires=n_wires)

# Генерация значений параметров для квантовых слоев
rand_params = np.random.uniform(high=2 * np.pi, size=(n_layers, n_wires))

@qml.qnode(dev)
def circuit(phi):
    # Кодирование 3 классических входных данных
    for j in range(n_wires):
        qml.RY(np.pi * phi[j], wires=j)

    # Случайная квантовая цепь
    RandomLayers(rand_params, wires=list(range(n_wires)))

    # Измерения, которые дают 3 классических выходных значений для следующих слоев
    return [qml.expval(qml.PauliZ(j)) for j in range(n_wires)]
```

Следующая функция `quanv` определяет квантовую свертку по следующей схеме:

- разделение изображения на блоки 2x2 пикселей;
- обработку каждого блока квантовой цепью `circuit`, описанной выше;
- 3 выходных значения для каждого блока помещаются в 3 разных канала одного пикселя выходного изображения.

Описанный выше процесс уменьшает разрешение входного изображения в 2 раза, что эквивалентно классической свертки с ядром 2x2 и шагом 2

```
def quanv(image):
    """Функция квантовой свертки над входным изображением."""

    out = np.zeros((14, 14, n_wires))

    # Циклы по координатам верхнего левого пикселя блоков 2x2
    for j in range(0, 28, 2):
        for k in range(0, 28, 2):
            # Обработка блока 2x2 из изображения квантовой цепью
            q_results = circuit(
                [
                    image[j, k, 0],
                    image[j, k + 1, 0],
                    image[j + 1, k, 0],
                    image[j + 1, k + 1, 0]
                ]
            )
            # Запись результатов наблюдения в выходной пиксель (j/2, k/2)
            for c in range(n_wires):
                out[j // 2, k // 2, c] = q_results[c]

    return out
```

Подготовка датасета

Так как квантовый сверточный слой в нашем случае не обучается, то лучше провести предварительную обработку им всех используемых изображений из датасета. Подготовленные данные будут использоваться классической нейронной сетью для обучения модели классификатора.

```
q_train_images = []
for idx, img in enumerate(train_images):
    q_train_images.append(quanv(img))
q_train_images = np.asarray(q_train_images)
print("Препроцессинг тренировочных изображений квантовой сверткой выполнен.")

q_test_images = []
for idx, img in enumerate(test_images):
    q_test_images.append(quanv(img))
q_test_images = np.asarray(q_test_images)
print("Препроцессинг тестовых изображений квантовой сверткой выполнен.")
```

Препроцессинг тренировочных изображений квантовой сверткой выполнен.

Препроцессинг тестовых изображений квантовой сверткой выполнен.

Давайте посмотрим на 4 первых тренировочных примера (первый ряд) и их 3 канала подготовленных данных (2-4 строки)

```

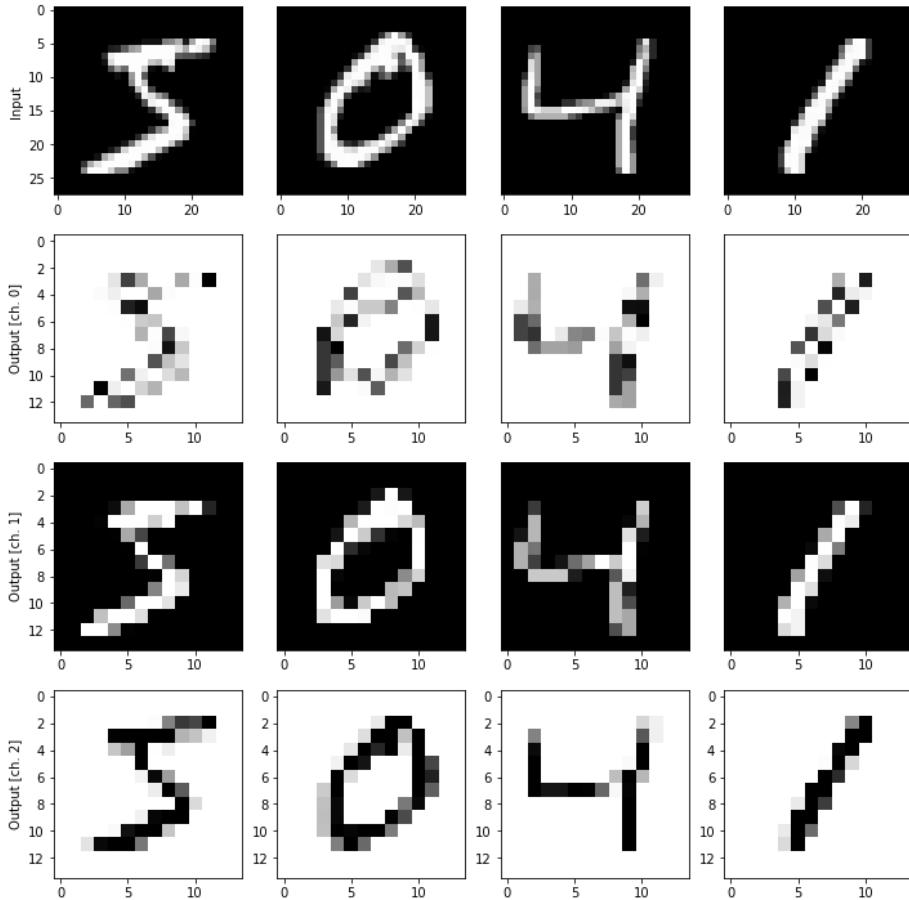
n_samples = 4
n_channels = n_wires
fig, axes = plt.subplots(1 + n_channels, n_samples, figsize=(10, 10))

for k in range(n_samples):
    axes[0, 0].set_ylabel("Input")
    if k != 0:
        axes[0, k].yaxis.set_visible(False)
        axes[0, k].imshow(train_images[k, :, :, 0], cmap="gray")

    # Отрисовка
    for c in range(n_channels):
        axes[c + 1, 0].set_ylabel(f"Output [ch. {c}]")
        if k != 0:
            axes[c, k].yaxis.set_visible(False)
            axes[c + 1, k].imshow(q_train_images[k, :, :, c], cmap="gray")

plt.tight_layout()
plt.show()

```



По обработанным данным (последние 3 ряда) видно уменьшение разрешения в 2 раза, а также то, что пространственная структура цифр сохраняется, что и должно наблюдаться для сверточных слоев

Гибридная квантово-классическая модель

Ниже опишем классическую часть нейронной сети, которая будет обучена для классификации 10 цифр. Будет использована очень простая модель, основанная на одном полно связном слое и финальном применении функции softmax, которая выдает вероятности представленных 10 классов.

Обучение будет осуществлено с помощью оптимизатора Adam по функции потерь, в качестве которой используется кросс-энтропия.

```

def MyModel():
    """Функция инициализирует и возвращает keras модель, готовая к обучению"""

    model = keras.models.Sequential([
        keras.layers.Flatten(),
        keras.layers.Dense(10, activation="softmax")
    ])

    model.compile(
        optimizer='adam',
        loss="sparse_categorical_crossentropy",
        metrics=["accuracy"],
    )
    return model

```

Тренировка моделей

Сначала обучим модель на данных, подготовленных с помощью квантовой цепи.

```

q_model = MyModel()

q_history = q_model.fit(
    q_train_images,
    train_labels,
    validation_data=(q_test_images, test_labels),
    batch_size=4,
    epochs=n_epochs,
    verbose=0,
)

```

Для сравнения с результатами чисто классической модели проведем ее обучение на входных необработанных картинках.

```

c_model = MyModel()

c_history = c_model.fit(
    train_images,
    train_labels,
    validation_data=(test_images, test_labels),
    batch_size=4,
    epochs=n_epochs,
    verbose=0,
)

```

Сравнение результатов

Сравнение проведем на основе тестовой точности и значению потерь в зависимости от номера эпохи

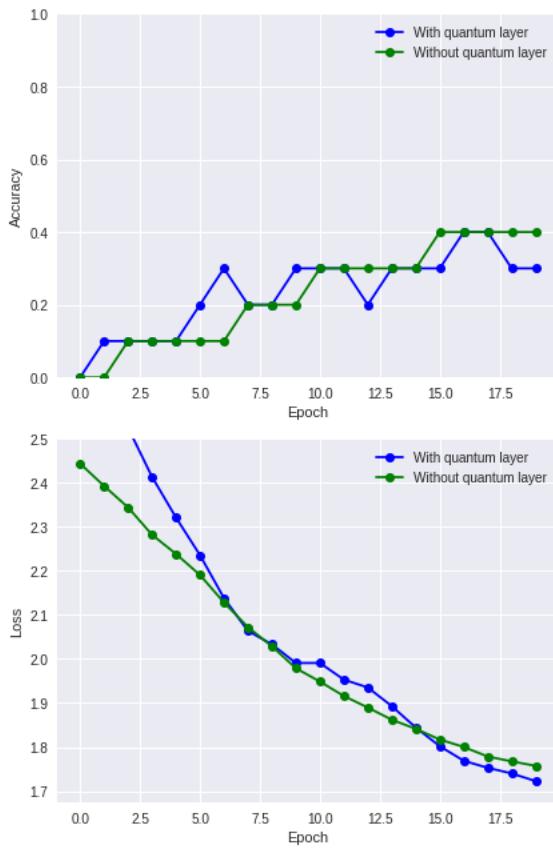
```

plt.style.use("seaborn")
fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(6, 9))

ax1.plot(q_history.history["val_accuracy"], "-ob", label="With quantum layer")
ax1.plot(c_history.history["val_accuracy"], "-og", label="Without quantum layer")
ax1.set_ylabel("Accuracy")
ax1.set_ylim([0, 1])
ax1.set_xlabel("Epoch")
ax1.legend()

ax2.plot(q_history.history["val_loss"], "-ob", label="With quantum layer")
ax2.plot(c_history.history["val_loss"], "-og", label="Without quantum layer")
ax2.set_ylabel("Loss")
ax2.set_ylim(top=2.5)
ax2.set_xlabel("Epoch")
ax2.legend()
plt.tight_layout()
plt.show()

```



Из рисунков видно наличие сильного сходства результатов для двух моделей: гибридной и чисто классической. В наших малых экспериментах достигнута довольно низкая точность в 40%. Однако при увеличении числа используемых примеров точность должна существенно выше.

Замечания

- в статье авторов подхода ([Henderson M. et.al., 2019](#)) для гибридной и чисто классической сети с большим количеством слоев (см. схему сети) и большим объемом данных, достигнута ожидаемая высокая точность выше 95%;
- блоки квантовой свертки сейчас особо интересны, так как они ориентированы на использование малого количества кубитов, которое доступно в текущий период времени устройств с промежуточным количеством кубитов NISQ.

Что мы узнали из лекции

- существует аналог сверточных нейронных сетей для квантовых вычислений;
- аналог квантовой свертки может быть реализован с использованием случайной квантовой цепи;
- использование квантовой свертки позволяет достичь результатов схожих с классическими сетями;
- основное преимущество квантовой свертки по отношению к классической в том, что первая является более широкой операцией, которая не может быть описана классическим случаем.

Квантовая химия. Введение.

Описание лекции

Из этой лекции мы узнаем:

- что такое квантовая химия, что с ее помощью можно сделать, а что – нельзя;
- как устроена самая простая теория, описывающая квантовую физику атома
- какие у этой теории есть ограничения

Введение

В предыдущих лекциях мы говорили о применении квантовой механики для вычислений и обработки информации. Однако исторически квантовая теория развивалась в первую очередь из-за того, что классическая физика не могла объяснить некоторые наблюдаемые эффекты, такие как дискретный спектр излучения атомов, фотоэффект в металлах, интерференцию частиц на дифракционных щелях.

Считается, что квантовая механика дает полное и точное описание состояния и эволюции любой системы при нерелятивистских условиях — по крайней мере, на сегодняшний день этому нет экспериментальных или теоретических противоречий. Это значит, что в теории возможно для любой системы частиц записать уравнения Шредингера, решить их и предсказать, как себя поведет система. Однако на практике оказывается, что в реальных задачах вроде моделирования лекарств и материалов просто “взять и посчитать” — задача весьма сложная, а иногда — неразрешимая.

Проблемами применения квантовой механики к химии и материаловедению занимается квантовая химия. Она делает это уже около 100 лет, по теме написаны толстые книжки с многоэтажными формулами, так что в лекции будут даны основы и простые примеры без полного вывода. Для желающих в конце приведены ссылки для углубленного изучения.

Предполагается, что читатель знаком с уравнением Шредингера и основными операторами (импульса, эволюции), бра-кет нотацией, а также помнит основы физики и химии на уровне старших классов школы.

АТОМ ВОДОРОДА

В чем проблема?

Спектр излучения и поглощения атомов — то есть на какой длине волн происходит поглощение и излучение света веществом — был одной из первых “нерешаемых” проблем, приведших в итоге к появлению квантовой физики. Для [простых веществ](#) в газообразной форме спектр является дискретным, и для атома водорода спектры поглощения и излучения в видимом диапазоне выглядят так:

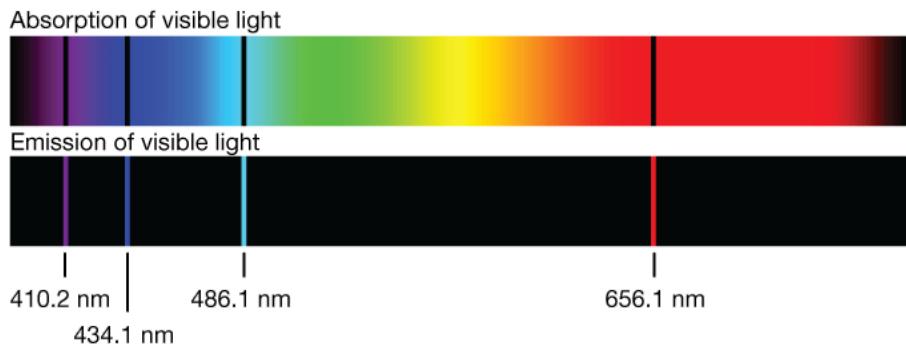


Fig. 113 Спектр поглощения и излучения водорода в видимом диапазоне

В конце 19 века было уже известно, что атом водорода состоит из двух заряженных частиц — протона и электрона, но из классической физики следовало, что спектр излучения такой системы должен быть непрерывным, что никак не стыковалось с наблюдениями (дискретные линии на графике). Попробуем вслед за исследователями квантовой физики начала 20 века разобраться с этой проблемой.

Note

Про имена. “Квантовая физика” и “квантовая механика” — связанные понятия, но не тождественные.

Квантовая механика — это теория (а точнее, множество теорий с различающимися терминологиями), в основе которой лежит аксиоматика о том, что сущности описываются волновыми функциями, что они эволюционируют и, зная эти функции, можно посчитать физические величины — например, энергию.

Квантовая физика — это область физики, исследующая квантовые эффекты, при этом она местами может не иметь строгого обоснования — лишь бы предсказания работали.

С квантовой химией тоже есть некоторая путаница — в двадцатом веке под ней имели в виду аналитические и численные методы решения задач квантовой механики применительно к молекулам и кристаллам. Но с развитием квантовых компьютеров эту область стали иногда называть вычислительной химией, а квантовой химией — применение квантовых компьютеров к этой области.

Быстрое и неправильное решение

Первой теорией в квантовой физике была [“Боровская теория”](#) – почти классическая теория, основанная на постуатах, разрешающих противоречия классической физики с экспериментами на уровне “хардкода”. Главный постулат заключается в том, что импульс может принимать только определенные дискретные значения (кванты), а не произвольные:

$$(p \cdot 2\pi r = n \cdot h)$$

$$(p = \frac{n \cdot h}{r})$$

где p – импульс, r – радиус орбиты электрона, h – константа Планка, n – любое целое число. Если применить этот постулат к атому водорода – паре из протона и электрона, то получится, что:

- электрон вокруг протона “летает по орбите”;
- спектр (уровни энергии электрона) – дискретный.

Можно показать это следующим образом.

Во-первых, поскольку частиц две и протон имеет массу много больше массы электрона, можно перейти в почти инерциальную систему отсчета протона, где он неподвижен, и решить задачу только для электрона.

Из классической механики берем [теорему о вириале](#) (“для стабильной системы из двух частиц, связанных потенциальными силами, в среднем кинетическая энергия равна половине потенциальной”):

$$(E_k = -E_p / 2)$$

$$(E = E_k + E_p = E_p / 2)$$

Записываем кинетическую энергию и потенциал Кулона для электрона в поле протона (в СИ):

$$(\frac{p^2}{2m} = k \frac{e^2}{2R}), \text{ где } k \text{ – постоянная из закона Кулона}$$

Используем постулат Бора и выразим “радиус” орбиты:

$$(\frac{n^2 \hbar^2}{2 m R^2} = k \frac{e^2}{2R})$$

$$(R = \frac{n^2 \hbar^2}{k m e^2})$$

Находим полную энергию:

$$(E = E_p / 2 = -\frac{k e^2}{2 R} = -\frac{k e^2}{2 \frac{n^2 \hbar^2}{k m e^2}} = -\frac{m e^4}{2 n^2 \hbar^2})$$

Уровни энергии пропорциональны $(\frac{1}{n^2})$ и именно [такая зависимость](#) наблюдается в эксперименте.

Посчитаем минимальную энергию ($n = 1$):

```
from scipy import constants as consts
k = (1 / (4 * consts.pi * consts.epsilon_0)) # постоянная кулонса
E1 = -k**2 * consts.m_e * consts.e**4 / (2 * consts.hbar ** 2) # энергия в Джоулях
E1_ev = E1 / consts.e # энергия в электрон-вольтах
print("Hydrogen Bohr ground state energy: {} eV")
```

```
Hydrogen Bohr ground state energy: -13.605693122885837 eV
```

Получившийся ответ 13.6 eV в точности совпадает с экспериментальным значением энергии ионизации водорода.

Итого: постулировав несколько очень удачных гипотез (главная – что импульс квантуется, то есть принимает только дискретные значения), удалось разрешить фундаментальную нестыковку между классической теорией и экспериментом: энергия стала дискретной, линии спектра стали обратно пропорциональны $\frac{1}{n}$, значения энергии ионизации совпадают с реальными измерениями.

Казалось бы, замечательно, проблема решена! Однако теория Бора имеет ряд проблем. Главная из них заключается в том, что теория работает только для “водородоподобных” атомов, то есть состоящих из ядра и одного электрона на внешней оболочке. Уже для атома гелия спектр не согласуется с боровской теорией.

Кроме того, с точки зрения науки теория Бора – это в некотором смысле “читечество”. Выбрав удачные постулаты и подставив их формулы в классическую физику, мы получили правильный результат. Но будь постулаты другими, результат получился бы тоже другой, так что фактически постулаты Бора – это гениальная догадка, позволившая угадать правильные результаты для некоторых систем.

Квантовая химия. Теория самосогласованного поля.

Описание лекции

Из этой лекции мы узнаем:

- как с помощью квантовой химии предсказать спектр атома водорода “из первых принципов”;
- как посчитать энергию атома с помощью Python;
- какие бывают волновые функции электронов и как их вычислять методом Self-Consistent Field;
- как посчитать энергию спирта.

Введение

Боровская теория оказалось недостаточной для описания таких сложных систем, как молекулы или не-водородоподобные атомы. В этой лекции мы разберемся с теорией самосогласованного поля – это один из первых вычислительных методов квантовой химии, пригодный на практике для моделирования разных систем (атомов, молекул, кристаллов).

Несмотря на то, что основа метода была разработана чуть ли не сто лет назад, он активно используется и сегодня, особенно с различными модификациями и дополнениями.

Что мы ищем?

С точки зрения квантовой химии, чаще всего мы хотим для произвольного набора частиц (атомов для молекулы или кристалла, протонов и электронов для атома) с известным потенциалом взаимодействия предсказывать стационарное состояние, т.е. находить стационарные волновые функции всех частиц системы. Квантовая механика постулирует, что для любой наблюдаемой физической величины существует оператор, которым мы можем действовать на волновую функцию и получить измеренное значение этой величины.

Например, если мы знаем волновые функции электронов и ядер в кристалле – можно предсказать теплопроводность кристалла. Или электрическую проводимость, или еще что-нибудь – поэтому основной задачей квантовой химии является поиск волновых функций-решений уравнения Шредингера для системы частиц.

Долгое и правильное решение, часть 1

Начнем с атома водорода и определения его уровней энергии – их можно сравнить с экспериментальными данными, а также результатами Боровской теории. Квантовая механика позволяет решить эту задачу, исходя только из “первых принципов”, то есть аксиоматики квантовой механики. При этом после некоторых усилий решения успешно обобщаются с атома водорода на любой другой.

Давайте опишем атом водорода — систему из электрона и протона — на языке квантовой механики, то есть в виде уравнения Шредингера, и посмотрим, что получится.

Электрон находится в потенциале протона, и его волновая функция должна зависеть от расстояния до протона.

Уравнение Шредингера для электрона будет следующим:

$$\left(\dot{\psi} + \frac{1}{\hbar^2} \frac{d}{dt} \langle \psi | \hat{H} | \psi \rangle \right) = -\frac{1}{2m} \nabla^2 \psi + V(r) \psi$$

Мы ищем решение, в котором электрон остается в атоме, то есть решаем стационарное уравнение, в котором $\langle \psi | \hat{H} | \psi \rangle$ не зависит от времени. Тогда оператор эволюции (в левой части уравнения Шредингера) при применении к $\langle \psi | \hat{H} | \psi \rangle$ должен вернуть нам тот же вектор $\langle \psi | \hat{H} | \psi \rangle$, умноженный на E, где E – это энергия частицы.

С точки зрения математики искомая волновая функция является собственным вектором оператора эволюции, а энергия – собственным значением.

С точки зрения объяснения “на пальцах” оператор эволюции при применении к волновой функции должен вернуть нам новую (эволюционированную/изменявшуюся во времени) волновую функцию. Если мы ищем стационарную — не меняющуюся во времени — волновую функцию, то при применении к ней оператора эволюции она не должна изменяться, иначе будет уже не стационарной.

$$(\dot{\langle imath \rangle} \hbar \frac{d}{dt} \langle \Psi(t, r) \rangle) \equiv \dot{\langle imath \rangle} \hbar \frac{d}{dt} \langle \Psi(r) \rangle \equiv E \{ \langle \Psi(r) \rangle \}$$

Оператор импульса раскрывается через градиент так: $\hat{p} = \dot{\langle imath \rangle} \hbar \nabla$

Потенциал взаимодействия двух частиц с противоположными единичными зарядами раскрывается так: $V(r) = -\frac{e^2}{r}$

Итого имеем:

$$(-\frac{\hbar^2}{2m} \nabla^2 - \frac{e^2}{r}) \langle \Psi(r) \rangle = E \{ \langle \Psi(r) \rangle \}$$

Прежде чем погружаться глубже в решение этого уравнения, нам надо разобраться с квантовыми числами.

Про квантовые числа

Теория Бора базируется на идее квантования импульса, и ее следствием является возникшее в формуле энергии число $\langle n \rangle$. Каждое число $\langle n \rangle$ соответствует определенному состоянию, в котором может находиться электрон и эти состояния отличаются энергией.

Идея о том, что электрон может находиться в одном состоянии из некоторого конечного набора вариантов, характеризуемого дискретными (квантовыми) числами, оказалась правильной. В дальнейшем она получила развитие в квантовой механике.

Число $\langle n \rangle$ называется главным [квантовым числом](#), оно определяет энергетический уровень электрона.

Из анализа свойств операторов гамильтонiana и момента импульса в квантовой механике выводятся еще три квантовых числа:

- $\langle l \rangle$ – орбитальное квантовое число, $0 \leq l \leq n-1$;
- $\langle m \rangle$ – магнитное квантовое число, $-l \leq m \leq l$;
- $\langle s \rangle$ – спиновое квантовое число, $s = 0|1\rangle$ (для атома водорода оно не играет роли, так как электрон только один).

В химии они [используются](#) в другой форме записи для описания орбиталей электронов в атоме.

Эти числа нам скоро понадобятся, чтобы описывать электрон в атоме.

Любая функция $\langle \Psi \rangle$, для которой уравнение выше верно, описывает электрон, который стабильно находится где-то около ядра и не покидает его, то есть образует с ним атом. При этом функций-решений уравнения на самом деле много, что физически соответствует тому, что электрон может находиться на разных орбиталах. Каждая орбита характеризуется квантовыми числами – $\langle n \rangle$, $\langle l \rangle$, $\langle m \rangle$ и $\langle s \rangle$, и обозначается как $\langle \Psi_{nlm} \rangle$. $\langle \Psi_{nlm} \rangle$ (орбита) с минимальной энергией E соответствует основному состоянию (ground state) – она описывает невозбужденный электрон. У водорода только один электрон, поэтому единственная $\langle \Psi_{nlm} \rangle$ с минимальной энергией соответствует невозбужденному атому водорода.

В целом основная задача квантовой химии – найти ground state произвольной системы частиц, так как это описывает “обычное” состояние системы, которое встречается чаще всего в реальности. Но если электрону в основном состоянии придать энергии, например, попасть в него другой частицей, то он может перейти на другую орбиталь, при этом оставшись частью атома. Со временем возбужденные атомы рано или поздно переходят в основное состояние, выбрасывая лишнюю энергию в виде фотона, что и дает спектр излучения атома.

Итого:

$\langle \Psi_{nlm} \rangle$ – это функция, соответствующая какой-то стабильной “траектории” (распределению плотности вероятности) электрона вокруг ядра, она является решением уравнения Шредингера, то есть собственной функцией гамильтонiana. Соответствующее ей собственное число – энергия электрона на этой орбитали. Для всех волновых функций с одним и тем же n энергия одинакова.

Если электрон описывается такой волновой функцией, то он часть атома, и если какой-либо электрон – часть атома, то он описывается такой волновой функцией. Электрон может переходить между этими волновыми функциями, получая и отдавая энергию, оставаясь при этом частью атома.

Физически электрон водорода обычно в $\langle \Psi | \Psi \rangle$, так как такое распределение соответствует минимальной энергии.

Долгое и правильное решение, часть 2

Note

А зачем нам вообще сдался спектр и энергии? Вслед за физиками двадцатого века нам приходится прорыться через дебри уравнений, абстракций и формул. Дело это непростое и может возникнуть резонный вопрос – а зачем оно все надо?

Помимо фундаментальной ценности в виде лучшего понимания устройства мира, объясняющая спектр теория дает крутое возможность: по измеренному спектру понять, что за вещество перед нами и какие процессы в нем протекают. Например, изучение спектров – один из немногих способов узнать что-то про процессы в звездах или биологических клетках. Для совсем практиков: спектроскопия также используется для детекции взрывчатых и наркотических веществ в малых дозах, а полевые транзисторы (основа большинства современных вычислительные устройства) работают на основе туннельного эффекта – квантового явления. Так что можно сказать, что квантовая механика используется сплошь и рядом!

Если мы перейдем в уравнении Х в сферические координаты со следующей параметризацией $\langle \Psi | \Psi \rangle = \frac{1}{r} Y(\theta, \phi)$ и воспользуемся несколькими волшебными выводами квантмеха [89], то получим:

$$\left(-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} \chi + \left(-\frac{e^2}{r} + \frac{\hbar^2}{2mr^2} (\ell(\ell+1)) \right) \chi = E \chi \right)$$

Все константы (массы, заряда, импульса, энергии) можно убрать, если перейти в кулоновские единицы измерений, где они приняты за единицу (то есть 1 единица заряда = заряд электрона, 1 единица массы = масса электрона).

$$\left(-\frac{d^2}{dr^2} \chi + \left(\ell(\ell+1) r^2 - \frac{1}{r} \right) \chi = E \chi \right)$$

Опустим несколько страниц выкладок [15a], учтем граничные условия и получим следующее решение:

$$R_{n\ell}(r) = r^\ell \cdot e^{-r/n} \cdot \sum_{k=0}^{n-\ell-1} \frac{(-2r/n)^k}{(2\ell+2+k)! (n-\ell-k-1)!} C$$

$$\langle \Psi | \Psi \rangle = R_{n\ell}(r) Y_{\ell m}(\theta, \phi),$$

где $Y_{\ell m}(\theta, \phi)$ – [сферические функции](#).

Если подставить это решение в уравнение Шредингера и найти энергию, то получим:

$$E_n = -\frac{1}{2m} \frac{e^4}{n^2 \hbar^2}$$

То есть получим ту же формулу, что и в теории Бора, и тот же численный результат – 13.6 eV.

Здесь начинает пропасть основная проблема квантовой химии – математическая и вычислительная сложность. Пока что проблема только концептуальная (сложно разобраться в формулах и уравнениях), но при росте числа частиц в системе даже отличное владение матаппаратом окажется недостаточным.

От теории к практике

Вернемся из начала двадцатого века обратно в день сегодняшний. Сто лет назад неберущиеся аналитически интегралы и замороченные дифференциальные уравнения были почти непреодолимыми препятствиями, поэтому поначалу теория квантовой химии развивалась в сторону более хитрых приближений и упрощений, позволяющих решить эти уравнения аналитически.

Сегодня реальные научные задачи решаются численно – и для этого написано множество высокопроизводительных пакетов, позволяющих на основе входных данных и ограничений вычислить определенным методом желаемые характеристики.

Как и в задачах машинного обучения, в первом приближении достаточно подготовить данные, скормить их волшебному комбайну-вычислителю и забрать ответ, но без понимания о происходящем под капотом есть шансы получить что-то неправильное.

В этой лекции мы воспользуемся Python-пакетом [psi4](#). Он реализует многие алгоритмы квантовой химии и имеет неплохое Python API.

Давайте посчитаем с помощью psi4 энергию основного состояния атома водорода. Некоторые параметры сейчас придется использовать, “поверив на слово”. Их смысл будет объяснен в дальнейшем.

```
import psi4
h_atom = psi4.geometry("H")
```

Мы задали атом водорода, по умолчанию атом помещается в начало координат.

```
psi4.set_options({
    'basis': 'STO-3G',
    'reference': 'rohf',
})
```

Тут уже поинтереснее – объяснение этих параметров пока отложим и вернемся после объяснения теории. В целом они определяют, каким именно методом и в каком базисе нужно численно решить уравнение Шредингера.

```
from scipy.constants import physical_constants
h2ev = physical_constants['hartree-electron volt relationship']

def e_in_ev(energy_in_ht):
    return energy_in_ht * h2ev[0]

e_in_ht = psi4.energy('scf') # энергия в единицах Hartree
print(f"Hydrogen ground state energy: {e_in_ev(e_in_ht)} eV")
```

Здесь мы посчитали энергию в единицах Хартри – специальной физической системе единиц, где истинная энергия атома водорода равна 1/2, и перевели ее в электрон-вольты.

Результат не очень точный (правильный, как мы помним, равен 13.6 eV) и мы его улучшим после того как разберемся с тем, что и как мы только что посчитали. Разобраться будет удобнее на примере атома гелия, потому что в атоме водорода есть только один электрон, а в любой реальной системе – больше одного.

Теория самосогласованного поля

Атом гелия

Следующим “по простоте” после атома водорода идет атом гелия – как говорит нам школьная химия, это атом из двух протонов, двух нейтронов и двух электронов. Протоны и нейтроны находятся близко друг к другу в ядре и имеют почти одинаковую массу, так что можно просто считать, что есть ядро с зарядом +2 и массой 4. А вот с электронами все сложнее: с одной стороны, это независимые частицы, а с другой – они взаимодействуют друг с другом по закону Кулона, так как оба имеют отрицательный заряд.

Note

Для гравитационного взаимодействия “[проблема трех тел](#)” не имеет известного аналитического решения.

Это означает, что если мы знаем, что где-то в глубоком космосе вдалеке от остального мира есть три объекта с известными массами, импульсами и координатами, то, увы, в общем случае не сможем предсказать их движение аналитически (хотя сможем предсказать численно, либо найти приближенное аналитическое решение, если масса одного объекта много больше других, например).

Для трех классических тел с кулоновским потенциалом все тоже сложно – можно посмотреть [тут](#), как поведет себя система трех тел с различными зарядами.

Попробуем записать уравнение Шредингера для системы из ядра и двух электронов:

$$\dot{\Psi} = \frac{1}{\hbar^2} \left(-\frac{1}{2m_e} \nabla_e^2 \Psi + \frac{e^2}{4\pi\epsilon_0 r_{AA}} \Psi + \frac{e^2}{4\pi\epsilon_0 r_{AB}} \Psi + \frac{e^2}{4\pi\epsilon_0 r_{BA}} \Psi \right)$$

Оно состоит из кинетической и потенциальной энергии двух электронов в поле ядра и (последнее слагаемое) потенциала взаимодействия двух электронов. Учитывая, что уравнение примерно такое же, как у водорода, только еще более сложное, решение стоит искать численно и с использованием приближений, а не аналитически.

Если бы не последний член гамильтониана, то можно было бы разбить все выражение на две независимых части – с переменными электрона A и с переменными электрона B. Так как это дифференциальное уравнение, можно было бы воспользоваться разделением переменных и найти отдельные решения для двух электронов – задача аналогична атому водорода, а ее мы уже решили.

Но из-за потенциала взаимодействия решение существенно усложняется, поскольку электроны влияют друг на друга. Придется прибегнуть к упрощениям – и одним из наиболее популярных подходов является *теория самосогласованного поля* (Self-Consistent Field).

Теория самосогласованного поля

Теория самосогласованного поля (self-consistent field theory) – это подход итеративного решения уравнения Шредингера для многочастичной системы, на основе которого построено много квантово-химических методов, наиболее известный из которых – метод Хартри-Фока. В снippetе выше строка `psi4.energy('scf')` означает, что энергия посчитана этим методом.

Основная идея теории заключается в следующем.

У нас есть несколько частиц, которые взаимодействуют между собой и найти цельное решение уравнения Шредингера для всех сразу не получается. Тогда вместо этого будем рассматривать частицы по очереди и считать, что все остальные действуют “в среднем” на выбранную частицу. То есть будем считать усредненный по пространству потенциал вместо точного.

Для электрона A в атоме гелия нам нужно учесть усредненное влияние электрона B. Можем для электрона B взять волновую функцию от атома водорода, посчитать на ее основе усредненное влияние на электрон A:

$$\langle (h_A + \hat{V}_{\text{eff}}) |\psi_A \rangle = \langle h_A + \frac{e^2}{r_{AB}} |\psi_B \rangle \langle \psi_B | \psi_A \rangle = \epsilon_A$$

Здесь $\langle h_A \rangle$ – это кинетическая энергия и потенциал ядра для электрона A, а $\langle V_{\text{eff}} \rangle$ – влияние электрона B на электрон A. $\langle V_{\text{eff}} \rangle$ вычисляется как влияние усредненной электронной плотности, распределенной в соответствии с $|\psi_B|$.

На этом шаге мы считаем, что $|\psi_B\rangle$ – фиксированная волновая функция и мы находим “переменную” $|\psi_A\rangle$. Однако $|\psi_A\rangle$ тоже будет влиять на $|\psi_B\rangle$ и, записав аналогичное уравнение для частицы B, мы следующим шагом найдем новую $|\psi_B\rangle$.

После изменения волновых функций каждого электрона мы можем заново записать их уравнения с новыми волновыми функциями и так по кругу, пока волновые функции и их энергии не сойдутся к какому-то стабильному (самосогласованному) решению.

Все вместе это создает итеративную процедуру:

1. На основе имеющихся волновых функций посчитать среднее поле, которое создают частицы (например, электрон B для электрона A).
2. Решить уравнение Шредингера с потенциалом, учитывающим среднее поле, то есть вычислить энергию и новые волновые функции.
3. Вернуться к шагу 1.

Для инициализации можно взять какие-то приблизительные волновые функции для всех частиц (двух электронов гелия), например, решения из уравнения водорода, то есть водородоподобные волновые функции. Далее итерацию повторяют, пока волновые функции и их энергии не перестанут изменяться, то есть самосогласуются.

Система таких уравнений, записанных для каждого электрона, называется **уравнениями Хартри**. На основе таких волновых функций для отдельных электронов можно собрать общую волновую функцию $|\Psi(r_A, r_B)\rangle$, самый простой вариант – это $|\Psi(r_A, r_B)\rangle = |\psi_A(r_A)\psi_B(r_B)\rangle$, он и был предложен первоначально. Однако есть проблема: такая $|\Psi\rangle$ получается не антисимметричной, а только такие волновые функции для системы электронов являются “физичными”, то есть могут существовать в реальности.

Note

Один из основополагающих принципов квантовой механики – тождественность частиц и их неразличимость. Принцип означает, что все частицы одного типа (например, электроны) одинаковы и характеризуются только своим состоянием. Например, если мы “переставим местами” два электрона в атоме (не только пространственно, но и в смысле их состояния и энергий), то получившийся атом будет неотличим от исходного.

Из принципа тождественности следует, что при перестановке не должна меняться плотность вероятности:

$$\langle |\Psi(X_A, X_B)|^2 = |\Psi(X_B, X_A)|^2 \rangle$$

Кроме того, представим что мы обменяли частицы дважды: А с В и обратно. Никакие физические свойства системы из-за этого измениться не должны. В общем случае, из этого не следует что волновая функция не изменилась – например, так как глобальная фаза волновой функции неизмерима, то мы могли получить $\langle|\Psi'|^2 = e^{i\theta} |\Psi|^2 \rangle$.

Но, если мы работаем больше чем в двух пространственных измерениях, то такой двойной обмен эквивалентен отсутствию обмена. Для начала представим что мы обмениваем частицы медленно описывая полукруг частицей А вокруг частицы В и потом сдвигая обе частицы. Тогда двойной обмен значит, что частица А описывает полный круг вокруг частицы В. Также, если мы непрерывно изменим ее маршрут не приближая ее к частице В, то мы ожидаем получить тот же результат.

Итак, какие замкнутые маршруты мы можем получить непрерывно деформируя маршрут в евклидовом пространстве без точки (В)? Ответ для любого измерения больше 2 – [любые](#). Следовательно двойной обмен должен давать такой же результат как и если бы мы просто оставили частицу А на месте [\[NSS+08\]](#).

У этого ограничения есть два решения: либо $\langle|\Psi(X_A, X_B)|^2 = |\Psi(X_B, X_A)|^2 \rangle$ (симметричность), либо $\langle|\Psi(X_A, X_B)|^2 = -|\Psi(X_B, X_A)|^2 \rangle$ (антисимметричность). У антисимметричных функций есть интересное свойство: если функция $f(x_1, x_2)$ антисимметрична, то $f(x_1=x, x_2=x)=0$, то есть антисимметричная функция равна нулю, если ее аргументы одинаковы. В этом легко убедиться на примере $f = x_1 - x_2$.

Для антисимметричной волновой функции это означает, что две частицы не могут иметь полностью одинаковое состояние – волновая функция (и вероятность) такой конфигурации равна нулю.

В эксперименте это строго выполняется для всех частиц с полуцелым спином (фермионов), к которым относятся электроны, а частицы с целым спином (бозоны), например, фотоны, имеют симметричную волновую функцию и могут иметь одинаковые состояния.

Для квантовой химии это все имеет одно важное следствие: электроны – это фермионы и волновая функция, описывающая всю систему электронов целиком, должна быть антисимметрична относительно перестановки (замены пары индексов). Любое не антисимметричное решение не физично, поскольку допускает существование электронов в одинаковом состоянии, а такого не бывает. В химии этот вывод называется [принцип запрета Паули](#).

Чтобы сделать волновую функцию системы из двух электронов антисимметричной, используется такой прием:

$$|\Psi(X_A, X_B)|^2 = |\psi_A(X_A)\psi_B(X_B) - \psi_B(X_A)\psi_A(X_B)|^2$$

Во-первых, легко проверить, что если поменять местами (X_A) и (X_B) , то вся функция просто изменит знак.

Во-вторых, можно заметить, что формулу можно записать как определитель матрицы: $\langle|\Psi(X_A, X_B)|^2 = \begin{vmatrix} \psi_A(X_A) & \psi_B(X_A) \\ \psi_A(X_B) & \psi_B(X_B) \end{vmatrix}$

Из курса линейной алгебры можно вспомнить, что определитель меняет знак при перестановке двух столбцов или двух строк – это свойство позволяет делать антисимметричные волновые функции систем из N волновых функций отдельных электронов, если использовать метод Хартри не для атома гелия, а для системы с большим числом электронов. Для этого составляется определитель $(N \times N)$ по аналогии с формулой выше: элемент в строке i, столбце j – это i-я волновая функция с параметрами j-го электрона в качестве аргумента.

Такой вариант сборки волновой функции системы частиц называется “определитель [Слэтера](#)”. Так как весь подход является аппроксимацией, не любая система может быть точно представлена таким детерминантом, но он является очень распространенным методом “сборки” волновой функции системы электронов в квантовой химии.

Его использование также немного меняет вид одноэлектронных уравнений: чтобы корректно учесть антисимметрию, в эффективный потенциал добавляется так называемое “обменное взаимодействие”.

Все вместе составляет метод Хартри-Фока:

- итеративная процедура самосогласованного поля;
- усредненное действие электронов друг на друга, учет обменного взаимодействия;
- детерминант Слетеера.

SCF в psi4

Теперь можно вернуться к коду и взглянуть на него чуть более осмысленно. При вычислении энергии мы явно передаем, что хотим посчитать ее методом Self-Consistent Field:

```
e_in_ht = psi4.energy('scf')
```

Но что происходит в настройках, пока по-прежнему неясно:

```
psi4.set_options({
    'basis': 'STO-3G',
    'reference': 'rohf',
})
```

Начнем с параметра `basis`. [STO-3G](#) – не стандарт связи, а **Slater Type Orbital** с **3 Гауссианами** в базисном наборе, то есть базис на основе детерминанта Слетеера. В описании метода SCF мы собирались начинать итерации с водородоподобных волновых функций, но так как весь метод является аппроксимирующим, нам никто не мешает выбрать другие волновые функции, если результаты лучше согласуются с экспериментом. Выбор базиса может существенно влиять на результат вычислений и современные базисы сложнее, чем Слетеоровский детерминант – он просто один из первых и наиболее популярных.

Параметр `reference` означает, какие предположения о волновой функции мы делаем, в данном случае используется [Restricted Open Shell Hartree-Fock](#), так как у атома водорода только один электрон и его оболочка не заполнена (на уровне энергии $n=1$ для этого нужно 2 электрона).

Давайте повторим вычисления с более “современными” опциями.

```
psi4.core.clean()

h_atom = psi4.geometry("H")

psi4.set_options({
    'basis': 'd-aug-cc-pv5z', # разбор этого базиса выходит за рамки этого интро
    'scf_type': 'pk',
    'reference': 'rohf'
})

e_in_ht = psi4.energy('scf')
print(f"Better hydrogen ground state energy: {e_in_ev(e_in_ht)} eV")
```

Та-дам! Используя более прокачанные базисы, мы получили правильный ответ.

Давайте посмотрим, что еще можно сделать с помощью self-consistent field.

Атом Гелия (численно)

Раз мы разобрали SCF на примере атома гелия, то наверняка можно посчитать его энергию в psi4.

```
psi4.core.clean()

he_atom = psi4.geometry("He")

psi4.set_options({
    'basis': 'STO-3G',
    'reference': 'rohf',
})

e_in_ht = psi4.energy('scf')
print(f"Helium ground state energy: {e_in_ev(e_in_ht)} eV")
```

Экспериментальное значение энергии атома гелия [равно](#) -79.0 eV.

Молекула водорода

Пока мы рассматривали только атомы, но SCF можно использовать и для молекул – потенциалы становятся сложнее, электронов больше, но общая логика не меняется.

```
psi4.core.clean()

h_mol = psi4.geometry("""
H 0 0 0
H 0 0 0.74
""") # задали 2 атома водорода с явными координатами

psi4.set_options({
    'basis': 'STO-3G',
    'reference': 'rohf',
})

e_in_ht_h = psi4.energy('scf', molecule=h_mol)
print(f"Hydrogen ground state energy: {e_in_ev(e_in_ht_h)} eV")
```

Здесь мы задали явно координаты обоих атомов водорода в молекуле и энергия электронов была вычислена в предположении, что ядра водородов неподвижны. Здесь расстояние в 0.74 Ангстрема взято из [экспериментальных данных](#). Если бы мы зададим неправильные координаты, то рассчитанная энергия окажется неверной. Точнее, она соответствовала бы нефизической ситуации, когда неведомая сила “удерживает” ядра водорода на месте.

В psi4 есть метод для оптимизации геометрии молекулы [psi4.optimize](#). Он не фиксирует положение ядер и возвращает минимальную возможную энергию с учетом вариации положения атомов.

Вычисления с оптимизацией геометрии занимают значительно больше времени.

```
psi4.core.clean()

h_mol_bad = psi4.geometry("""
H 0 0 0
H 0 0 1.5
""") # неверное расстояние в ангстремах

psi4.set_options({
    'basis': 'STO-3G',
    'reference': 'rohf',
})

# рассчитываем энергию "в точке" с неправильной геометрией
e_in_ht_h_bad = psi4.energy('scf', molecule=h_mol_bad)

# рассчитываем энергию, оптимизируя по ходу геометрии
e_in_ht_h_optimized = psi4.optimize('scf', molecule=h_mol_bad)

print(f"Hydrogen molecule, incorrect ground state energy: {e_in_ev(e_in_ht_h_bad)} eV")
print(f"Hydrogen molecule, optimized ground state energy:
{e_in_ev(e_in_ht_h_optimized)} eV")
```

Для некорректной геометрии получилась завышенная энергия, а после оптимизации – почти что такая же энергию, как при вычислении с фиксированным расстоянием 0.74. В оптимальном состоянии энергия системы должна быть минимальна, так что результаты вполне разумны.

Подобный метод можно использовать и для поиска геометрии куда более сложных молекул.

Молекула этилового спирта

Молекула водорода – это все еще почти игрушечный пример. Давайте попробуем обсчитать молекулу этанола.

Задавать руками геометрию молекулы C2H5OH можно, но будет явно сложнее, чем для молекулы водорода. К счастью, это необязательно: psi4 умеет скачивать геометрию из базы данных [PubChem](#) по номенклатурному имени либо уникальному ChemId.

```
psi4.core.clean()

eth = psi4.geometry("pubchem:ethanol")

psi4.set_options({
    'basis': 'STO-3G',
    'reference': 'rohf',
})

e_in_ht_eth = psi4.energy('scf', molecule=eth)

print(f"Ethanol ground state energy: {e_in_ev(e_in_ht_eth)} eV")
```

Итоги

Мы разобрались с базовой теорией квантовой химии:

- как записать уравнение Шредингера для атома;
- какое получается аналитическое решение для атома водорода;
- как устроен метод Self-Consistent Field для вычисления волновых функций и энергии для задачи многих тел;
- как пользоваться SCF в python пакете psi4.

В примерах мы везде вычисляли ground state энергию, но, конечно, зная волновые функции, можно посчитать многое еще. Например, можно вычислить спектр поглощения и энергию ионизации (энергии возбужденных состояний), моделировать взаимодействие молекул (найти равновесное состояние для двух систем), с помощью плагинов можно смоделировать рассеяния рентгена на молекуле... Квантовая механика постулирует, что любая измеримая величина является усреднением определенного оператора по волновой функции, поэтому возможности ограничены в основном вычислительной сложностью, а не теорией.

Для более глубокого погружения в практику квантовой химии можно пройти лабораторные работы psi4: [раз](#), [два](#).

Модель Изинга

В этой лекции мы познакомимся с моделью Изинга, которая изначально была разработана для описания магнетизма, но оказалась настолько удачной и универсальной, что сегодня к решению именно этой задачи пытаются свести многие проблемы реального мира, причем не только из физики. В следующем блоке мы подробно покажем, как к гамильтонианам типа Изинга, или, по-другому, "спиновым стеклам" могут быть сведены задачи комбинаторной оптимизации и квантовой химии. Так что знакомство с этой удивительной моделью, а также описывающим ее гамильтонианом нам просто необходимо!

Note

Специальные квантовые компьютеры компании D-Wave сконструированы так, что они могут решать вообще только одну задачу – нахождения основного состояния гамильтонианов типа Изинга. Но эта задача настолько распространена и важна, что эти компьютеры стали первыми в мире коммерческими квантовыми компьютерами! Кстати, далее этим компьютерам у нас посвящена [отдельная лекция](#).

Далее мы посвятим довольно много времени объяснению этой модели. Это может показаться скучным и занудным, но это важно для понимания того, как это все работает и как решать с помощью вариационных квантовых алгоритмов реальные задачи!

Задача Изинга в одномерном случае

Note

Ниже мы попробуем на пальцах объяснить модель Изинга. Пробовать мы будем через цепочку атомов антиферромагнетика во внешнем магнитном поле. Если вы плохо помните физику и вам это объяснение покажется сложным, то не расстраивайтесь – дальше мы также объясним задачу Изинга как задачу о поиске максимального разреза в графе – известную задачу комбинаторной оптимизации.

Пусть у нас есть, например, цепочка атомов, которые обладают магнитным моментом. Например, цепочка атомов антиферромагнетика. И мы прикладываем к этой цепочке внешнее магнитное поле.

Тогда, если поле маленькое, наши атомы будут стараться выстроиться в антиферромагнитный порядок, когда соседние из них имеют моменты, направленные в разные стороны. Но если поле уже большое, то оно будет стремиться “поворнуть” моменты по своему направлению. А если еще вспомнить, что магнитный момент атома является квантовой величиной и может быть в суперпозиции состояний в одну сторону и в противоположную, то не очень маленькое, но и не слишком большое поле будет переводить часть атомов именно в такие суперпозиции.

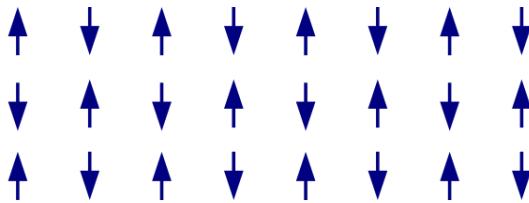


Fig. 114 Иллюстрация антиферромагнитного порядка

Note

Reminder о квантовой физике В квантовой механике есть фундаментальное уравнение, которое описывает динамику квантовых систем. Оно называется уравнением Шредингера: $\imath\hbar\frac{\partial|\Psi\rangle}{\partial t} = \hat{H}|\Psi\rangle$, где \hat{H} – это оператор Гамильтона, или гамильтониан. Также его называют оператором полной энергии системы, так как в общем случае он равен сумме операторов кинетической и потенциальной энергии. Начиная с этой лекции мы будем очень часто обращаться к этому оператору, но в целом в нем нет ничего принципиально сложного. Это такой же эрмитов оператор, как и другие. А наблюдаемая величина, которую мы получаем при [измерении](#) этого оператора – это энергия системы.

Давайте теперь запишем гамильтониан такой системы. Для представления магнитных моментов мы будем использовать оператор (σ_z) – другими словами, спин в направлении оси (Z) . Если кто-то забыл, как выглядит оператор (σ_z) , то рекомендуем еще раз просмотреть [раздел про операторы Паули первой лекции](#). Далее мы будем очень активно использовать эти матрицы для представления задач реального мира!

Для начала, в случае если внешнего поля нет, мы должны записать взаимодействие соседних атомов. Так как у нас антиферромагнетик, минимальная энергия достигается в случае, если у нас каждый спин противонаправлен с соседними. Это просто оператор $(\sigma_z_j\sigma_z_{j+1})$, который действует на все пары соседних спинов. Ну и сразу введем некоторую константу обменного взаимодействия (J) , чтобы потом нам было удобно сравнивать ее с внешним полем. В итоге, для цепочки из (N) спинов, получаем:

$$\hat{H}_{h=0} = J \sum_{i=0}^{N-1} (\sigma_z_i\sigma_z_{i+1})$$

А теперь давайте добавим внешнее поле (h) . В этом случае поле просто действует на все спины и пытается выстроить их в зависимости от своего направления, например, вниз. Тогда полный гамильтониан такой системы можно записать в виде:

$$\hat{H}_{h \neq 0} = J \sum_{i=0}^{N-1} (\hat{\sigma}_z_i\hat{\sigma}_z_{i+1} - h(\sigma_z_i))$$

Задача Изинга как задача о максимальном разрезе в графе

Задача о максимальном разрезе в графе – это очень известная задача комбинаторики. Она относится к классу (NP) -трудных, и к ней можно свести все другие (NP) задачи. При этом ее формулировка одна из самых простых среди всего класса задач. Формулируется она следующим образом.

Нам дан граф – набор вершин (V) и связывающих их ребер (E) . Нам надо найти такое разделение вершин (V) на два непересекающихся набора (V_1, V_2) , что число ребер между вершинами из разных наборов будет максимально.

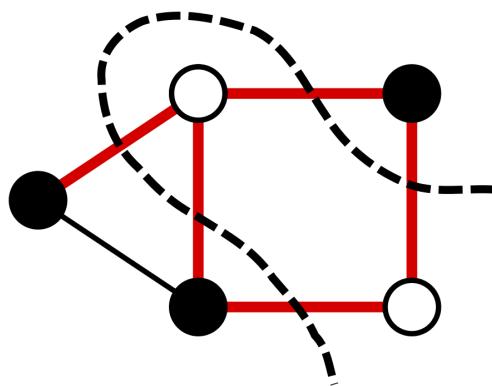


Fig. 115 Иллюстрация задачи о максимальном разрезе в графе

Теперь давайте представим, что каждой вершине нашего графа мы сопоставили кубит. Для этих кубитов мы можем [производить измерения](#) по оси $\langle Z \rangle$, чтобы понять, как направлен тот или иной спин. И давайте запишем вот такой гамильтониан и внимательно на него посмотрим:

$$\hat{H} = \sum_{\langle u,v \rangle \in E} \hat{\sigma}_z^u \hat{\sigma}_z^v$$

Тут суммирование $\langle u,v \rangle \in E$ идет по всем ребрам графа, а $\langle u,v \rangle$ – вершины инцидентные ребром. Если вспомнить, что собственные значения $\langle \sigma_z \rangle$ это $\langle \pm 1 \rangle$ для, соответственно, спина “вверх” и спина “вниз”, то не трудно понять, в каком случае у нас будет минимум энергии этого гамильтониана. А будет он тогда, когда максимальное число пар вершин $\langle u,v \rangle$ имеют разную ориентацию своих спинов. Ведь если они имеют одинаковую направленность (причем не важно, $\langle +1 \rangle$ или $\langle -1 \rangle$), их произведение будет равно $\langle 1 \rangle$, но если направленность разная, то их произведение даст нам $\langle -1 \rangle$. Таким образом, минимум энергии такого гамильтониана достигается тогда, когда мы разбили наши вершины на две группы – спин “вверх” и спин “вниз” – причем разбили так, что число ребер между этими группами максимальное. А это в чистом виде формулировка задачи о максимальном разрезе в графе!

Note

Тематика квантовой физики мало обсуждалась в первых лекциях, но нам пока достаточно знать лишь то, что для любой физической системы, включая квантовую, стремится в состояние с минимальной энергией. Например, тело, подброшенное вверх, стремится упасть на землю, а возбужденный атом стремится релаксировать в невозбужденное состояние.

При этом из квантовой физики мы знаем, что для реальных физических систем наиболее вероятными являются состояния с минимальной энергией и системы стремятся в эти состояния прийти. Теперь для простоты предположим, что наш график – это просто цепочка, то есть ребра есть лишь между соседними в одномерном пространстве вершинами. Ну и теперь давайте сформулируем нашу задачу о максимальном разрезе чуточку сложнее – нам надо найти не просто максимальный разрез, а такой разрез, который самый большой при наименьшем числе вершин в наборе $\langle V_1 \rangle$. И поскольку теперь у нас два вклада в стоимость, то нам нужны коэффициенты, которые покажут, что нам важнее. Пусть это будут $\langle J \rangle$ и $\langle h \rangle$. Тогда гамильтониан соответствующей модели Изинга можно записать так:

$$H = J \sum_{i=0}^{N-1} \hat{\sigma}_z^i \hat{\sigma}_z^{i+1} - h \sum_{i=0}^N \hat{\sigma}_z^i$$

Как видно, это тот же самый гамильтониан, который мы получили и для моделирования антиферромагнетиков. То есть задача об основном состоянии цепочки антиферромагнитных частиц во внешнем поле эквивалентна задаче о максимальном разрезе в графике-цепочке при некотором штрафе за одно из выделенных направлений спинов.

Эквивалентность в данном случае значит, что:

- решив задачу о максимальном разрезе, мы найдем основное состояние физической системы
- как-то смоделировав физическую систему, подождав пока она релаксирует, после чего измерив ее, мы получим конфигурацию, отвечающую решению задачи о максимальном разрезе

Note

Одномерная цепочка атомов, или поиск максимального разреза в графе-цепочке, является простым случаем и не является NP-задачей. Однако уже в двумерном случае эта задача становится сильно сложнее, как и, например, если в цепочке атомов ферромагнетика мы добавим взаимодействие не только соседних спинов, но и взаимодействие с соседями соседа. Аналогично, модель вида Изинга сильно усложняется при добавлении недиагональных (off-diagonal elements) элементов гамильтониана, например, когда внешнее поле направлено в другом направлении и второй член гамильтониана принимает вид $\sum_{i=1}^N \sigma_i^x$. Более подробное исследование данной модели приводится в [этой продвинутой лекции](#).

Модель Изинга на чистом NumPy

Давайте попробуем реализовать одномерный гамильтониан Изинга на чистом [NumPy/SciPy](#) в виде разреженной матрицы. Для этого вспомним, что действуя оператором σ_z на i -й кубит, мы одновременно действуем единичным оператором на все остальные, а потом перемножаем все операторы произведением Кронекера. Из [лекций по линейной алгебре](#) мы помним также об ассоциативности произведения Кронекера, чем и воспользуемся:

```
import numpy as np
from scipy import sparse
from scipy.sparse import linalg as sl

def sigmaz_k(k: int, n: int) -> sparse.csr_matrix:
    left_part = sparse.eye(2 ** k)
    right_part = sparse.eye(2 ** (n - 1 - k))

    return sparse.kron(
        sparse.kron(
            left_part,
            sparse.csr_matrix(np.array([[1, 0], [0, -1]])))
        ),
        right_part
    )
```

А теперь мы можем реализовать и сам оператор Изинга:

```
def ising(j: float, h: float, n: int) -> sparse.csr_matrix:
    res = sparse.csr_matrix((2 ** n, 2 ** n), dtype=np.complex64)

    for i in range(n - 1):
        res += j * sigmaz_k(i, n) * sigmaz_k(i + 1, n)
        res -= h * sigmaz_k(i, n)

    res -= h * sigmaz_k(n - 1, n)

    return res
```

Если внешнего поля нет, спины выстраиваются в полный антиферромагнитный порядок, в чем легко убедиться. Создадим оператор для такой модели и, например, 10 спинов (или 10 вершин в графе, если мы говорим в терминах Max-Cut):

```
op = ising(1, 0, 10)
solution = sl.eigs(op, which="SR", k=1, return_eigenvectors=True)
print(f"Energy: {solution[0][0]}")
```

```
Energy: (-9.000000000000018-2.208468631860285e-16j)
```

Note

Тут мы пользуемся функциями из [ARPACK](#) – набором рутин для линейной алгебры разреженных систем. Более подробно о способах и алгоритмах классических решений задачи о собственных значениях мы расскажем в одной из следующих лекций, полностью посвященной этой теме. Пока же мы просто используем эту рутину как “черный ящик”. Более подробное описание этой функции и ее аргументов можно посмотреть в [документации библиотеки SciPy](#).

Эта энергия соответствует антиферромагнитному порядку, в этом легко убедиться, нарисовав спины и формулу на бумажке. Внимательный читатель заметил, что в этот раз мы вернули также и первый собственный вектор, который в нашем случае является волновой функцией основного состояния. А как мы помним, квадраты элементов вектора волновой функции дают нам вероятности соответствующих битовых строк (если для вас это все звучит дико, то очень рекомендуем вернуться к [лекции про кубит](#)). Давайте посмотрим на эту битовую строку, или, другими словами, на порядок наших спинов в решении (или на разбиение вершин графа на два подмножества в терминах Max-Cut):

```
def probs2bit_str(probs: np.array) -> str:
    size = int(np.log2(probs.shape[0]))
    bit_s_num = np.where(probs == probs.max())[0][0]
    s = f"{bit_s_num:b}"
    s = "0" * (size - len(s)) + s

    return s

probs = solution[1] * solution[1].conj()
print(probs2bit_str(probs))
```

0101010101

Теперь давайте попробуем добавить внешнее поле с коэффициентом, равным удвоенному значению константы обменного взаимодействия. В терминах комбинаторной задачи, мы добавляем штраф, равный два умножить на число спинов, направленных вверх.

```
op = ising(1, 2, 10)
solution = sl.eigs(op, which="SR", k=1, return_eigenvectors=True)
print(f"Energy: {solution[0][0]}")

probs = solution[1] * solution[1].conj()
print(probs2bit_str(probs))
```

Energy: (-10.99999999999999-1.0571225056728771e-16j)
0101010010

Видим, что теперь наш антиферромагнитный порядок уже не полный. В целом, данная модель довольно интересная, так как при некотором отношении $\frac{h}{J}$ у нас происходит фазовый переход от полной упорядоченности, а при дальнейшем росте h мы приходим к одинаковой ориентации всех спинов, в чем легко убедиться, взяв, например, $h = 100$:

```
op = ising(1, 100, 10)
solution = sl.eigs(op, which="SR", k=1, return_eigenvectors=True)
print(f"Energy: {solution[0][0]}")

probs = solution[1] * solution[1].conj()
print(probs2bit_str(probs))
```

Energy: (-991.000000000039-3.4882612571137096e-14j)
0000000000

Заключение

В этой лекции мы на базовом уровне познакомились с моделью Изинга – очень важным концептом в квантовом машинном обучении. Мы узнали, что:

- Модель Изинга изначально была создана для объяснения магнетизма
- Нахождение решений для модели Изинга в общем случае – *NP-полная* задача
- Модель Изинга также может быть сформулирована в терминах задачи о максимальном разрезе в графе (и наоборот)
- В классической модели Изинга существуют интересные фазовые переходы
- Модель Изинга легко реализовать в коде, используя *SciPy*, но размерность задачи растет очень быстро

Точное аналитическое решение одномерной модели Изинга.
Метод трансфер-матрицы.

В этой лекции мы более подробно обсудим модель Изинга, [рассмотренную ранее](#). Мы чуть глубже погрузимся в физику этой модели, а также рассмотрим ее частные случаи, которые имеют точные аналитические решения. Для этого мы также познакомимся с методом трансфер-матрицы.

Модель Изинга

Модель Изинга представляет собой широко распространенную математическую модель статистической физики, которая может применяться в различных областях человеческой деятельности.

Note

Не смотря на свою известность многие не знают, что модель названа в честь Эрнеста Изинга – автора работы, опубликованной в 1925 г. [Isi25], в которой он впервые представил данную модель и которая стала его первой и по сути единственной публикацией.



Fig. 116 Эрнст Изинг, 1900-1998

Суть самой модели можно постичь, например, при стремлении описать свойства намагничивания материала. Представим себе, что у нас имеется магнитная решетка (она может быть одномерной, двумерной или трехмерной) каждой вершине которой сопоставляется число называемое спином ($\langle\sigma\rangle$) и принимающее значение $\langle+1\rangle$ или $\langle-1\rangle$ (соответственно поле « $\langle\uparrow\rangle$ » или поле « $\langle\downarrow\rangle$ »). Каждому из $\langle 2^N \rangle$ (где N – число вершин решётки) возможных вариантов расположения спинов (или намагниченности) приписывается энергия, величина которой определяется условиями попарного взаимодействия спинов соседних вершин.

Например, энергия одномерной модели Изинга, в которой спины взаимодействуют только с ближайшими соседями (посредством параметра обменного взаимодействия $\langle J \rangle$) и с однородным внешним полем $\langle H \rangle$, имеет вид:

$$\langle E(\langle\sigma\rangle) \rangle = -J \sum_{i=1}^N \sigma_i \sigma_{i+1} - H \sum_{i=1}^N \sigma_i. \quad (18)$$

Это выражение соответствует периодическим граничным условиям, при которых первый и $\langle N+1 \rangle$ -й спины отождествляются, то есть

$$\langle \sigma_{N+1} \rangle \equiv \sigma_1.$$

В этом случае можно считать, что система спинов находится на окружности, то есть первый и $\langle N \rangle$ -й спины являются ближайшими соседями и взаимодействуют между собой. Для описания такой системы необходимо найти статистическую сумму $\langle Z \rangle$, в которой содержится полная информация о всех состояниях системы. Тогда другие характеристики системы, такие как удельная энергия, свободная энергия, энтропия, намагниченность, магнитная восприимчивость и др. вычисляются через статистическую сумму или ее производные.

Статистическая сумма рассматриваемой модели при температуре $\langle \theta \rangle$ есть сумма по всем спиновым конфигурациям $\langle \{\sigma\} \rangle$

$$\langle Z(\theta, H, N) \rangle = \sum_{\{\sigma\}} e^{-\frac{E(\{\sigma\})}{\theta}}. \quad (19)$$

Сумму по состояниям $\langle \{\sigma\} \rangle$ можно понимать как последовательное суммирование по значениям всех переменных $\langle \sigma_i \rangle$:

$$\langle \sum_{\{\sigma\}} \rangle \equiv \sum_{\sigma_1 = \pm 1} \sum_{\sigma_2 = \pm 1} \dots \sum_{\sigma_N = \pm 1}.$$

Следует заметить, что количество членов статистической суммы экспоненциально растет с увеличением числа вершин, соответственно увеличивается и время расчета термодинамических параметров для большого количества частиц.

Матричный подход

Для аналитического решения модели Изинга как правило применяется матричный подход. В рамках такого подхода статистическая сумма многих решеточных моделей Изинга может быть записана в виде следа некоторой степени *трансфер-матрицы*. Чтобы выразить статистическую сумму (19) через трансфер-матрицу, сначала представим гиббсовскую экспоненту в виде произведения $\lambda(N)$ множителей

$$\begin{aligned} \left[e^{-\frac{E}{kT}} \right] = T_{\sigma_1} T_{\sigma_2} T_{\sigma_3} \dots T_{\sigma_{N-1}} T_{\sigma_N} \\ T_{\sigma_N} T_{\sigma_1} = \prod_{j=1}^N T_{\sigma_j} T_{\sigma_{j+1}} \end{aligned} \quad (20)$$

где

$$\langle T_{\{\sigma\}} \rangle = \langle \sigma | T | \sigma' \rangle = e^{\frac{h}{2}(\sigma + \sigma')}, \quad (21)$$

а $\backslash(t)$ и $\backslash(h)$ введены для более компактного написания формул и выражаются через $\backslash(J)$, $\backslash(H)$ и $\backslash(theta)$

следующим образом

$$\quad t = \frac{J}{H} \theta, \quad h = \frac{H}{J} \theta. \quad]$$

Числа $\{\sigma\}$ будем дальше рассматривать как матричные элементы матрицы $\{T\}$, строки и столбцы которой вместо натуральных чисел нумеруются изинговскими переменными $\{\sigma\}$ и $\{\sigma'\}$. Так как изинговские переменные принимают два значения: $\{\sigma, \sigma' = \pm 1\}$, то матрица $\{T\}$ является квадратной матрицей размера $\{2 \times 2\}$. Договоримся дальше считать, что значение $\{\sigma = 1\}$ ($\{\sigma' = 1\}$) соответствует первой строке (первому столбцу), а значение $\{\sigma = -1\}$ ($\{\sigma' = -1\}$) соответствует второй строке (второму столбцу). Матрица $\{T\}$ с элементами (21) в таком случае имеет следующий явный вид:

$\| \begin{aligned} & \text{\begin{matrix} T = \begin{pmatrix} T_{\{1,1\}} & T_{\{1,-1\}} \\ T_{\{-1,1\}} & T_{\{-1,-1\}} \end{pmatrix} \end{matrix}} \\ & \text{\begin{matrix} T_{\{1,1\}} & T_{\{1,-1\}} \\ T_{\{-1,1\}} & T_{\{-1,-1\}} \end{pmatrix} } \end{aligned} \|$

Матрица (V^T) называется *трансфер-матрицей Крамерса–Ваннье*.

Равенство (20) позволяет записать статистическую сумму (19) в следующем виде

$$\langle Z(\theta, H, N) \rangle = \sum_{\{\sigma_1 = \pm 1\}} \sum_{\{\sigma_2 = \pm 1\}} \dots \sum_{\{\sigma_{N-1} = \pm 1\}} \sum_{\{\sigma_N = \pm 1\}} T_{\{\sigma_1 \sigma_2\}} T_{\{\sigma_2 \sigma_3\}} \dots T_{\{\sigma_{N-1} \sigma_N\}} T_{\{\sigma_N \sigma_1\}} \quad (23)$$

Структура слагаемых поз

$$\sum \sigma_1 = \pm 1 \sum \sigma_2 = \pm 1 \dots \sum \sigma_{N-1} = \pm 1 \sum \sigma_N = \pm 1$$

$$T_{\sigma_1} \begin{aligned} &= \sum_{\sigma_2} T_{\sigma_1 \sigma_2} \sum_{\sigma_3} T_{\sigma_1 \sigma_2 \sigma_3} \dots T_{\sigma_1 \sigma_{N-1} \sigma_N} T_{\sigma_1 \sigma_N \sigma_1} = N \\ &\quad \vdots \\ &= \sum_{\sigma_1} T_{\sigma_1} \sum_{\sigma_2} T_{\sigma_1 \sigma_2} \sum_{\sigma_3} T_{\sigma_1 \sigma_2 \sigma_3} \dots T_{\sigma_1 \sigma_{N-1} \sigma_N} T_{\sigma_1 \sigma_N \sigma_1} = N \end{aligned} \quad (24)$$

Теперь, поскольку $(T_{\{\sigma\} \sigma'})$ являются матричными элементами матрицы (T) , заметим, что сумма

$$[\sum_{\sigma_N = \pm 1} T_{\{\sigma_{N-1}\}} \sigma_N T_{\{\sigma_N \sigma_1\}}]$$

равна $(\prod_{i=1}^{N-1} \sigma_i)^2$, то есть матричному элементу матрицы (I^2) , которая получается в результате (матричного) умножения двух матриц $\langle T \rangle$ (является квадратом матрицы $\langle T \rangle$). Тогда сумма в правой части равенства (24), равна

$$\begin{aligned} & \left[\sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} \dots \sum_{\sigma_{N-1}=\pm 1} T_{\sigma_1 \sigma_2} T_{\sigma_2 \sigma_3} \dots T_{\sigma_{N-2} \sigma_{N-1}} T^{\sigma_{N-1} \sigma_1} \right] \quad (25) \\ & \sigma_3 \dots T_{\sigma_{N-2} \sigma_{N-1}} T^{\sigma_{N-1} \sigma_1} \right] \end{aligned}$$

Дальше используем равенство

$$\nabla \cdot (\sum_{n=1}^N \sigma_{n-1} T_n \sigma_{n-2} \sigma_{n-1}) T^2 \cdot (\sigma_{N-1} \sigma_1) = T^3 \cdot (\sigma_{N-2} \sigma_1) \nabla$$

где $(T^3 = T \cdot T^2)$ – куб матрицы (T) , и перепишем выражение (25) в виде

```
\[ \sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} \dots \sum_{\sigma_{N-2}=\pm 1} T_{\sigma_1 \sigma_2} T_{\sigma_2 \sigma_3} \dots T_{\sigma_{N-3} \sigma_{N-2}} T^{N-2} \sigma_{N-2} \sigma_1. \]
```

Продолжим этот процесс и просуммируем последовательно по всем $\{\sigma_i\}$, $(i = 2, \dots, N)$, кроме $\{\sigma_1\}$, тогда равенство (23) преобразуется к виду

$$\{ Z(\theta, H, N) = \sum_{\sigma_1=\pm 1} T^N_{\sigma_1} \}$$

В последнем равенстве суммируются диагональные элементы матрицы (T^N) , которая есть матрица

- Формула (26) является следствием трансляционной инвариантности рассматриваемой системы. Трансляционная инвариантность в данном случае означает, что энергия системы спинов не изменяется при пространственном сдвиге, то есть спиновые конфигурации $\{\sigma\}$ и $\{\sigma'\}$, где $\sigma_i = \sigma_{i+1}$, имеют одинаковые энергии.
 - Представление гиббсовской экспоненты в виде произведения (20), где множители $T(\sigma \sigma')$ имеют вид (21), не является единственным возможным. Можно выбрать трансфер-матрицу по-другому. Например, если взять $\tilde{T}(\sigma \sigma') = e^{(\sigma \sigma' + h\sigma)}$, то тогда: $e^{-\frac{E(\sigma)}{kT}} = \prod_i T(\sigma_i \sigma_{i+1})$. Таким образом, в выборе трансфер-матрицы существует некоторый произвол. Значение статистической суммы не зависит от этого произвола, поэтому им естественно распорядиться исходя из каких-то дополнительных соображений, например так,

чтобы матрица обладала какой-то симметрией, что упрощает вычисления. Для рассмотренной выше одномерной модели Изинга трансфер-матрица была выбрана вещественной симметричной (то есть эрмитовой).

3. Если выбранная нами трансфер-матрица $\langle T \rangle$ является эрмитовой, то с помощью унитарного преобразования $\langle U \rangle$ ее можно преобразовать к диагональной форме:

$$\langle \grave{T} \rangle = U^{-1} T U = dgn\{\lambda_1, \lambda_2, \dots, \lambda_r\},$$

где символом $\langle dgn \rangle$ обозначена диагональная матрица, на главной диагонали которой расположены ее собственные значения. В общем случае размерность такой матрицы равна величине $\langle r=s^n \rangle$, где $\langle s \rangle$ — число состояний спина в вершине, а $\langle n \rangle$ — число обменных взаимодействий спинов цепочки (в рассматриваемой нами одномерной модели Изинга $\langle s=2 \rangle$ и $\langle n=1 \rangle$). В этом случае статистическая сумма (26) переписывается к следующему виду:

$$\langle Z(\theta, H, N) = \langle \text{Tr} \rangle, \grave{T}^N = \sum_i \langle r \rangle \lambda_i^N, \rangle \quad (27)$$

При этом величина статистической суммы не меняется, поскольку след матрицы является ее инвариантом.

Отметим, что процедуру диагонализации трансфер-матрицы можно провести множеством способов. В зависимости от подбора диагонализирующей матрицы $\langle U \rangle$ её можно провести таким образом, что на первом месте в матрице $\langle \grave{T} \rangle$ будет стоять её главное (единственное максимальное вещественное) собственное значение, которое всегда существует по теореме Фробениуса–Перрона. Если на первое место в спектре трансфер-матрицы поставить ее главное собственное значение, то статистическую сумму (27) можно переписать как:

$$\langle Z(\theta, H, N) = \lambda_1^N \left[1 + \frac{\lambda_2}{\lambda_1} + \dots + \frac{\lambda_r}{\lambda_1} \right]^N \quad (28)$$

Тогда при переходе к термодинамическому пределу $\langle (N \rightarrow \infty) \rangle$ в выражении {eqn:11} все слагаемые, кроме первого, стремятся к нулю,

$\langle \begin{aligned} & \lim_{N \rightarrow \infty} \left(\frac{\lambda_2}{\lambda_1} \right)^N = 0, \dots, \lim_{N \rightarrow \infty} \left(\frac{\lambda_r}{\lambda_1} \right)^N = 0 \end{aligned} \rangle$

и выражение для статистической суммы преобразуется к виду: $\langle Z = \lambda_1^N \rangle$.

В результате все термодинамические и магнитные функции системы – свободная энергия системы, приходящаяся на один спин,

$$\langle F = -\frac{\partial \langle Z \rangle}{\partial \theta} = -\theta \lambda_1 \rangle \quad (30)$$

энтропия

$$\langle S = -\frac{\partial F}{\partial \theta} = \ln \lambda_1 + \frac{\lambda_1}{\lambda_1 - 1} \rangle \quad (31)$$

теплоемкость

$$\langle C = -\frac{\partial^2 F}{\partial \theta^2} = 2 \frac{\lambda_1}{\lambda_1 - 1} \left(1 + \frac{\lambda_1}{\lambda_1 - 1} \right) \rangle \quad (32)$$

намагниченность

$$\langle M = -\frac{\partial F}{\partial H} = \frac{\lambda_1}{\lambda_1 - 1} \rangle \quad (33)$$

и магнитная восприимчивость

$$\langle \chi = -\frac{\partial^2 F}{\partial H^2} = \frac{\lambda_1^2}{\lambda_1 - 1} \rangle \quad (34)$$

определяются только через главное собственное значение трансфер-матрицы.

Решение одномерной модели

Теперь вернемся к аналитическому решению рассматриваемой нами классической одномерной модели Изинга ($s=2$) и ($n=1$). Как мы показали задача сводится к сведению трансфер-матрицы (22) диагональному виду. Для этого воспользуемся правилами линейной алгебры и решим характеристическое уравнение для трансфер-матрицы и найдем её собственные значения:

$$\langle \begin{aligned} \det(T - \lambda I) &= \det(T - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) \\ &= 0, \end{aligned} \rangle \quad (35)$$

где $\langle I \rangle$ – единичная матрица, $\langle \lambda \rangle$ – действительная переменная. Раскрывая определитель (35), находим искомое характеристическое уравнение:

$$\langle \lambda^2 - \lambda (e^{t+h} + e^{t-h}) + e^{2t} - e^{-2t} = 0 \rangle \quad (36)$$

решением которого являются два собственных значения:

$$\langle \begin{aligned} \lambda_{\pm} &= \frac{1}{2} (e^{t+h} + e^{t-h}) \pm \sqrt{\frac{1}{4} (e^{t+h} + e^{t-h})^2 - e^{2t} + e^{-2t}} \\ &= e^t \left(\cosh(h) \pm \sqrt{\cosh^2(h) - \sinh^2(h)} \right), \end{aligned} \rangle \quad (37)$$

где $\langle \cosh(h) = \sqrt{e^h + e^{-h}} \rangle$ – гиперболический косинус и $\langle \sinh(h) = \sqrt{e^h - e^{-h}} \rangle$ – гиперболический синус. Следовательно статистическая сумма рассматриваемой модели записывается в следующем виде: $\langle Z = \text{Tr } T^N \rangle = \lambda_+^N + \lambda_-^N$. Здесь следует отметить, что для всех физических значений параметров $\langle t \rangle$ и $\langle h \rangle$ ($\langle \ln \mathbb{R} \rangle$, $\langle \infty \rangle$) выполняется условие: $\langle |\lambda_+| > |\lambda_-| \rangle$. Тогда в термодинамическом пределе ($\langle N \rightarrow \infty \rangle$) значение $\langle \lambda_+ \rangle$ преобладает над $\langle \lambda_- \rangle$ и

удовлетворяет (29), а статистическая сумма преобразуется к виду: $\langle Z \rangle = \lambda_+^N + \lambda_-^N \sim$
 $\lambda_+^N \langle N \rangle$). Отсюда можно получить выражение для удельной свободной энергии системы, приходящейся
на один спин (30) (с точностью до множителя $(-\theta)$):

$$f(t,h) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left(\lambda_+^N \lambda_-^N \right) = t + \ln \left(\cosh(h) + \sqrt{\sinh^2(h) + e^{-4t}} \right)$$

и для намагнченности (33):

$$m = -\frac{\partial f(t,h)}{\partial h} = \frac{\sinh(h) + \sinh(h)\cosh(h)}{\sqrt{\sinh^2(h) + e^{-4t}}} \left(\cosh(h) + \sqrt{\sinh^2(h) + e^{-4t}} \right)$$

которая является аналитической функцией $|h|$ и исчезает при $|h| \rightarrow 0$ (так как $\langle \sinh(h) \rangle \rightarrow 0$). Таким образом рассматриваемая модель не допускает фазового перехода при любой положительной температуре ($\theta > 0$). Она претерпевает *парамагнитный-ферромагнитный* фазовый переход только при $\theta = 0$, в этом случае спонтанная намагнченность равна $|m| = 1$.

Заключение

В этой лекции мы познакомились с аналитическим решением одномерной модели Изинга методом трансфер-матрицы. В заключении хотелось бы сказать, что модель Изинга поддается многим обобщениям, будь то более высокие измерения, различные решетки или модифицированные взаимодействия.

О блоке “D-Wave”

Этот блок будет посвящен квантовому отжигателю компании D-Wave и задачам, которые можно решать с его помощью.

D-Wave: квантовый отжиг, оптимизация и семплинг

Описание лекции

В этой лекции мы рассмотрим квантовый отжигатель компании **D-Wave** и задачи, которые можно решать с его помощью. Мы заострим основное внимание на физике и алгоритмах, нежели на hardware-реализации этого устройства. Эта лекция основана на [официальной документации D-Wave](#).

Введение в квантовый отжиг

На сегодняшний день, квантовый отжиг представляет собой отдельную модель квантовых вычислений, не связанную с так называемыми **Gate-Model** квантовыми вычислениями. Эта парадигма позволяет эффективно решать задачи комбинаторной оптимизации больших масштабов, представляющие большой интерес для многих областей науки и индустрии. Вкратце, в модели квантового отжига на физическую модель (коей является массив кубитов) отображаются параметры решаемой задачи, после чего кубиты адиабатически (достаточно медленно) переводятся в основное состояние, которое и кодирует в себе искомое решение. Давайте теперь разберем подробнее данный подход.

Задачи оптимизации и семплирования

Квантовый отжигатель (**Quantum Annealing Processor**) в силу своей физической реализации приводит к решениям, соответствующим низкоэнергетическим состояниям физической системы. В то время как задачи оптимизации (**Optimization Problems**) требуют глобального минимума энергии, иногда хватает и семплов из состояний, близких к основным. Такие задачи называются задачами семплирования (**Probabilistic Sampling Problems**).

Классическим примером задачи оптимизации в контексте квантового отжига является, например, [задача коммивояжера](#). Квантовая физика, в силу своей природы, может помочь решать такого рода задачи, так как они могут быть переформулированы в задачи дискретной минимизации энергии. Одно из основных правил физики состоит в том, что все физические системы стремятся к состояниям с минимальной энергией. Например, объекты скользят с холмов, горячие вещи со временем остывают и так далее. Этот принцип соблюдается и в мире квантовой физики. Квантовый отжиг просто использует квантовую физику для поиска низкоэнергетических состояний системы и, как следствие, оптимального или почти оптимального решения.

В свою очередь, задачи семплирования, часто являющиеся вычислительно сложными (например, задача семплирования из ограниченной машины Больцмана), идеально подходят к квантовым отжигателям в случае когда необходимо построить вероятностную модель, а доступ к полному вероятностному распределению состояний не представляется возможным. Такая постановка задачи подходит например, чтобы обучить модель генерировать изображения рукописных цифр из знаменитого датасета [MNIST](#). По этой теме можно послушать [доклад](#) Никиты Щуцкого “Quantum-Classical Neural Networks in Computational Drug Design”, сделанный в рамках трека ODS по квантовым вычислениям.

Квантовая физика под капотом

Как мы уже знаем, квантовый гамильтониан является оператором, ставящим в соответствие собственным состояниям системы ее энергии. Множество собственных чисел гамильтониана называется его спектром. В квантовом отжигателе [D-Wave](#) гамильтониан имеет следующую форму:

$$\mathcal{H}_{Ising} = \underbrace{\{-A(t)\left(\sum_i \hat{\sigma}_x^i\right)^2\}}_{\text{Initial Hamiltonian}} + \underbrace{\{B(t)\left(\sum_i h_i \hat{\sigma}_z^i\right) + \sum_{i,j} J_{i,j} \hat{\sigma}_z^i \hat{\sigma}_z^j\}}_{\text{Final Hamiltonian}}$$

, где $A(t)$, $B(t)$ являются функциями от времени $t \in [0, T]$, $\hat{\sigma}_x^i$ – операторами Паули, действующими на кубиты, а h_i и $J_{i,j}$ – локальными полями и взаимодействиями кубитов соответственно. Физически, вектор $\langle h \rangle$ соответствует внешнему магнитному полю, применимому к каждому кубиту. В свою очередь, матрица $\langle J \rangle$ отвечает за попарные $\langle ZZ \rangle$ взаимодействия кубитов. Функции $A(t)$, $B(t)$ должны быть такими, что $A(0)=B(T)=1$, $A(T)=B(0)=0$, то есть гамильтониан со временем полностью переходит из $\langle H_{Initial} \rangle$ в $\langle H_{Final} \rangle$. Вид этих функций называется расписанием отжига или [annealing schedule](#).

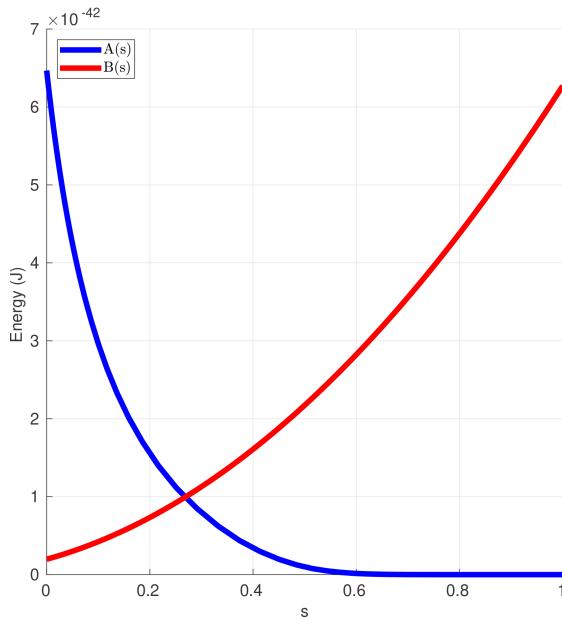


Fig. 117 Пример расписания отжига: функций $A(t)$, $B(t)$.

Отметим также, что матрица взаимодействий кубитов $\langle J_{i,j} \rangle$ ограничена графом взаимодействий кубитов, который обычно фиксирован. Для практического применения в коде детальное представление этого графа знать не нужно, но ознакомиться с ним можно на специально посвященной этому [странице документации D-Wave](#).

Разберем подробно, что же представляет собой данный гамильтониан. Начальный гамильтониан $\langle H_{Initial} \rangle$ представляет собой гамильтониан с основным состоянием состоящим из суперпозиции $|0\rangle$ и $|1\rangle$. Этот гамильтониан в литературе также называется [tunneling Hamiltonian](#). Финальный гамильтониан $\langle H_{Final} \rangle$, в свою очередь, является гамильтонианом задачи (его называют также [Problem Hamiltonian](#)), чье основное состояние и представляет собой решение, параметры которого находятся в матрице взаимодействий кубитов $\langle J_{i,j} \rangle$ и векторе локальных полей $\langle h_i \rangle$. Обратим наше внимание на тот факт, что $\langle H_{Final} \rangle$ – полностью классический гамильтониан, он диагонален в вычислительном базисе (базисе $\langle \sigma_z \rangle$). Каждое собственное (классическое) состояние этого гамильтониана соответствует классической функции потерь для решаемой задачи.

Таким образом, в квантовом отжиге система инициализируется в основное состояние начального гамильтониана – такое состояние легко приготовить экспериментально, то есть система легко переходит в основное состояние начального гамильтониана в силу его простоты. По мере самого отжига, в силу [адиабатической теоремы](#), основное состояние начального гамильтониана переходит в основное состояние финального гамильтониана. Для наглядности, на графике ниже (взят из [этой](#) статьи) показано типичное поведение первых энергетических уровней гамильтониана. Основному состоянию соответствует кривая с наименьшей энергией.

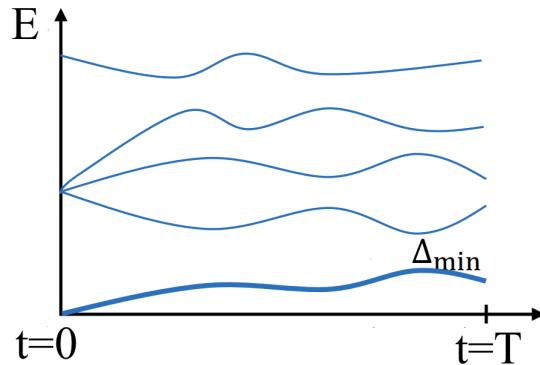


Fig. 118 Типичная зависимость от времени энергетических уровней гамильтонианов, используемых в квантовом отжиге

Очень важно отметить, что для успеха данного алгоритма необходимо выполнение условий адиабатической теоремы:

1. во-первых, минимальная энергетическая щель (Δ_{\min}) (разница энергий основного и первого возбужденного состояний) не должна обращаться в 0 в любой момент времени,
2. а во-вторых, функции $(A(t))$ и $(B(t))$ должны меняться достаточно медленно.

Можно показать, что второе условие может быть переформулировано в виде $(T \gg \frac{1}{\Delta_{\min}})$. Именно это неравенство ответственно за наибольшую проблему, с которой можно столкнуться при практической реализации данного алгоритма. В частности, из него видно, что время алгоритма зависит от минимальной (за все время алгоритма) энергетической щели, которая не должна быть слишком маленькой, иначе время алгоритма станет слишком большим. Если же щель будет слишком большой (что соответствует быстрому отжигу), то система может перейти в возбужденное состояние. Однако, основные состояния начального и финального гамильтонианов являются, качественно, сильно разными. Основное состояние начального гамильтониана является тривиальным, в то время как основное состояние финального гамильтониана кодирует в себе решение сложной задачи оптимизации. Это наталкивает нас на мысль, что в течение процесса квантового отжига наша система может пережить квантовый фазовый переход.

Для квантовых фазовых переходов первого рода известно, что $(\Delta_{\min} \propto e^{-N})$, где N – количество кубитов, что приводит к $(T \propto e^{2N})$, то есть экспоненциально большому времени алгоритма. В этом случае задача считается “сложной”, так как алгоритм не масштабируется.

В свою очередь, для квантовых фазовых переходов второго рода верно, что $(\Delta_{\min} \propto N^{-l})$ ($(l > 0)$), что приводит к $(T \propto N^{2l})$, то есть полиномиальному времени алгоритма и задача считается “простой”.

Пример: задача раскраски графа

Давайте теперь попробуем решить с помощью квантового отжига следующую задачу комбинаторной оптимизации – [задачу раскраски графа](#), или, как она называется в англоязычной литературе – [graph coloring problem](#). В этой задаче нам необходимо с помощью заданного числа цветов присвоить цвет каждой вершине графа так, чтобы любые две вершины, имеющие общее ребро, имели разные цвета. Назовем наш граф $(G=(E,V))$, где E – множество ребер, а V – множество вершин. Для наглядности, мы рассмотрим эту задачу в виде раскраски карты Канады четырьмя цветами – нам необходимо каждой канадской провинции (аналогу штата в США) присвоить цвет так, чтобы все граничащие провинции имели разные цвета. Этот пример находится также в [документации D-Wave](#). Визуально, решение этой задачи будет выглядеть как на следующем графике.



Fig. 119 Раскрас карты Канады с помощью четырех цветов.

D-Wave предоставляют нам доступ к трем видам бэкендов: квантовому, классическому и гибридному. Так как наш курс посвящен квантовому машинному обучению, мы воспользуемся непосредственно квантовым отжигателем. Для этого нам необходимо получить токен, который генерируется в личном кабинете пользователя при регистрации в [облачном сервисе D-Wave](#).

Импортируем нужные библиотеки.

```
import dwavebinarycsp
from dwave.system import DWaveSampler, EmbeddingComposite
import networkx as nx
import matplotlib.pyplot as plt
```

Отобразим нашу задачу на графике, где провинции будут вершинами графа, а границы – его ребрами. Например, (AB, BC) – ребро, соответствующее границе между провинциями Альберта и Британская Колумбия. Обозначения провинций представлены в таблице ниже.

Code	Province	Code	Province
AB	Alberta	BC	British Columbia
MB	Manitoba	NB	New Brunswick
NL	Newfoundland and Labrador NS		Nova Scotia
NT	Northwest Territories	NU	Nunavut
ON	Ontario	PE	Prince Edward Island
QC	Quebec	SK	Saskatchewan
YT	Yukon		

Обозначения провинций Канады

```
provinces = ['AB', 'BC', 'MB', 'NB', 'NL', 'NS', 'NT', 'NU', 'ON', 'PE', 'QC', 'SK',
'YT']
neighbors = [('AB', 'BC'), ('AB', 'NT'), ('AB', 'SK'), ('BC', 'NT'), ('BC', 'YT'),
('MB', 'NU'), ('MB', 'ON'), ('MB', 'SK'), ('NB', 'NS'), ('NB', 'QC'),
('NL', 'QC'), ('NT', 'NU'), ('NT', 'SK'), ('NT', 'YT'), ('ON', 'QC')]
```

Чтобы “объяснить” нашему отжигателю эту задачу, нам необходимо сформулировать ее в виде [задачи удовлетворения ограничений](#), или, как она называется в англоязычной литературе, – [Constraint Satisfaction Problem \(CSP\)](#). В свою очередь, библиотеки D-Wave умеют отображать ее на модель Изинга. Математически, это

выглядит следующим образом: каждой провинции (вершине графа) сопоставляется 4 бинарные переменные (каждая отвечает за один из четырех цветов) $\{x_i \mid i \in \{0,1\}\}$, где одна из переменных “выбирает” цвет и переключается в состояние $\{x_i=1\}$, в свою очередь остальные три переменные остаются в состоянии $\{0\}$.

Теперь займемся ограничениями. Следующая функция выражает ограничение того, что соседние вершины в графе не могут быть одного и того же цвета. Область ее определения и значения – булевы переменные.

```
def not_both_1(v: bool, u: bool) -> bool:  
    return not (v and u)
```

Зададим наши конфигурации цветов.

```
one_color_configurations = {(0, 0, 0, 1), (0, 0, 1, 0), (0, 1, 0, 0), (1, 0, 0, 0)}  
colors = len(one_color_configurations)
```

Создадим объект класса **CSP** и зададим следующие ограничения: каждая провинция имеет лишь один цвет и каждая ей соседняя не может иметь такой же цвет.

```
csp = dwavebinarycsp.ConstraintSatisfactionProblem(dwavebinarycsp.BINARY)  
  
for province in provinces:  
    variables = [province+str(i) for i in range(colors)]  
    csp.add_constraint(one_color_configurations, variables)  
  
for neighbor in neighbors:  
    v, u = neighbor  
    for i in range(colors):  
        variables = [v+str(i), u+str(i)]  
        csp.add_constraint(not_both_1, variables)
```

Конвертируем нашу модель из **CSP** в **BQM (Binary Quadratic Model)** – бинарную квадратичную модель. Это модель Изинга, в которой значения спинов $\{\sigma_i = 0.5 * (x_{i+1})\}$. Таким образом, каждая провинция Канады и ее цвет будет кодироваться четырьмя спинами, один из которых направлен вверх (вниз), пока остальные направлены вниз (вверх).

```
bqm = dwavebinarycsp.stitch(csp)
```

Создадим наш семплер, передав в него **API Token**:

```
import os  
base_sampler = DWaveSampler(token=os.environ["DWAVE_TOKEN"])
```

Дадим задание нашему квантовому семплеру с количеством семплов $\{1000\}$.

```
sampler = EmbeddingComposite(base_sampler)  
sampleset = sampler.sample(bqm, num_reads=1000, label='Canada Map Coloring')
```

Проверим, что было найдено решение задачи,

```
sample = sampleset.first.sample  
if not csp.check(sample):  
    print("Failed to color map. Try sampling again.")  
else:  
    print(sample)
```

```
{'AB0': 1, 'AB1': 0, 'AB2': 0, 'AB3': 0, 'BC0': 0, 'BC1': 0, 'BC2': 1, 'BC3': 0,  
'MB0': 1, 'MB1': 0, 'MB2': 0, 'MB3': 0, 'NB0': 0, 'NB1': 0, 'NB2': 0, 'NB3': 1,  
'NL0': 0, 'NL1': 0, 'NL2': 0, 'NL3': 1, 'NS0': 1, 'NS1': 0, 'NS2': 0, 'NS3': 0,  
'NT0': 0, 'NT1': 1, 'NT2': 0, 'NT3': 0, 'NU0': 0, 'NU1': 0, 'NU2': 1, 'NU3': 0,  
'ON0': 0, 'ON1': 0, 'ON2': 1, 'ON3': 0, 'PE0': 0, 'PE1': 0, 'PE2': 0, 'PE3': 1,  
'QC0': 1, 'QC1': 0, 'QC2': 0, 'QC3': 0, 'SK0': 0, 'SK1': 0, 'SK2': 1, 'SK3': 0,  
'YT0': 0, 'YT1': 0, 'YT2': 0, 'YT3': 1}
```

Мы получили словарь **sample**, который будет выглядеть следующим образом: его ключи будут кодировать провинцию и цвет, а значения – бинарные переменные, означающие раскрашена ли провинция этим цветом. Этот словарь – семпл из состояния, полученного при решении данной задачи методом квантового отжига. Построим визуализацию нашего решения в виде графа с раскрашенными вершинами.

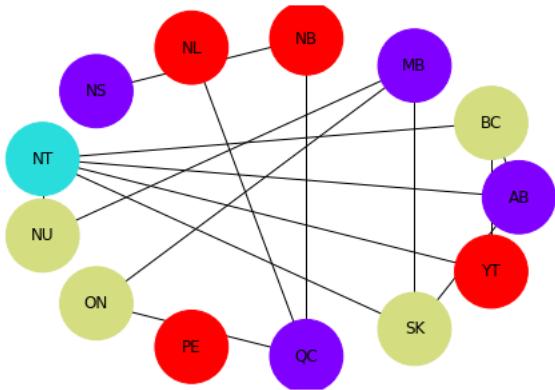
```

from typing import List
from typing import Tuple
from typing import Dict

def plot_map(sample: Dict[str, bool], provinces: List[str], neighbors:
List[Tuple[str]]):
    G = nx.Graph()
    G.add_nodes_from(provinces)
    G.add_edges_from(neighbors)
    color_map = {}
    for province in provinces:
        for i in range(len(colors)):
            if sample[province+str(i)]:
                color_map[province] = i
    node_colors = [color_map.get(node) for node in G.nodes()]
    nx.draw_circular(G, with_labels=True, node_color=node_colors, node_size=3000,
cmap=plt.cm.rainbow)
    plt.show()

plot_map(sample, provinces, neighbors)

```



После запуска функции `plot_map`, мы получим изображение графа, в котором вершины будут соответствовать провинциям и цветам, а ребра – границам между провинциями. Вы можете сами убедиться, что ни одна провинция не будет иметь такой же цвет, как и ее сосед. Также отметим, что такая задача может иметь не одно решение, поэтому не нужно ожидать, что граф будет получаться всегда одним и тем же.

Таким образом, мы сумели решить задачу раскраски графа с помощью квантового отжигателя.

Список литературы

[ion]

Single trapped atom captures science photography competition's top prize. Accessed: 2021-10-05. URL:

<https://webarchive.nationalarchives.gov.uk/ukgwa/20200930161403/https://epsrc.ukri.org/newsevents/news/single-trapped-atom-captures-science-photography-competitions-top-prize/>.

[Mon]

Varenna summer school on ion traps: quantum simulation, lecture i. Accessed: 2021-10-05. URL:

https://iontrap.umd.edu/wp-content/uploads/2014/01/Lecture1_QSIM.pptx.

[AMunozNK21]

Shahnawaz Ahmed, Carlos Sánchez Muñoz, Franco Nori, and Anton Frisk Kockum. Quantum state tomography with conditional generative adversarial networks. *Physical Review Letters*, 127(14):140502, 2021.

[doi:https://doi.org/10.1103/PhysRevLett.127.140502](https://doi.org/10.1103/PhysRevLett.127.140502).

[And19]

Neculai Andrei. A diagonal quasi-newton updating method for unconstrained optimization. *Numerical Algorithms*, 81(2):575–590, 2019. [doi:https://doi.org/10.1007/s11075-018-0562-7](https://doi.org/10.1007/s11075-018-0562-7).

[ABBradler+21]

J. M. Arrazola, V. Bergholm, K. Brádler, T. R. Bromley, M. J. Collins, I. Dhand, A. Fumagalli, T. Gerrits, A. Goussev, L. G. Helt, J. Hundal, T. Isacsson, R. B. Israel, J. Izaac, S. Jahangiri, R. Janik, N. Killoran, S. P. Kumar, J. Lavoie, A. E. Lita, D. H. Mahler, M. Menotti, B. Morrison, S. W. Nam, L. Neuhaus, H. Y. Qi, N. Quesada, A. Repington, K. K.

Sabapathy, M. Schuld, D. Su, J. Swinarton, A. Száva, K. Tan, P. Tan, V. D. Vaidya, Z. Vernon, Z. Zabaneh, and Y. Zhang. Quantum circuits with many photons on a programmable nanophotonic chip. *Nature*, 591(7848):54–60, mar 2021. [arXiv:2103.02109](https://arxiv.org/abs/2103.02109), doi:[10.1038/s41586-021-03202-1](https://doi.org/10.1038/s41586-021-03202-1).

[AAB+19]

Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, and others. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019. doi:<https://doi.org/10.1038/s41586-019-1666-5>.

[BAG21]

Afrad Basheer, A. Afham, and Sandeep K. Goyal. Quantum k -nearest neighbors algorithm. 2021. [arXiv:2003.09187](https://arxiv.org/abs/2003.09187).

[BV04]

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004. ISBN 0521833787. URL: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike-20&path=ASIN/0521833787>.

[BVM+21]

Michael Broughton, Guillaume Verdon, Trevor McCourt, Antonio J. Martinez, Jae Hyeon Yoo, Sergei V. Isakov, Philip Massey, Ramin Halavati, Murphy Yuezhen Niu, Alexander Zlokapa, Evan Peters, Owen Lockwood, Andrea Skolik, Sofiene Jerbi, Vedran Dunjko, Martin Leib, Michael Streif, David Von Dollen, Hongxiang Chen, Shuxiang Cao, Roeland Wiersema, Hsin-Yuan Huang, Jarrod R. McClean, Ryan Babbush, Sergio Boixo, Dave Bacon, Alan K. Ho, Hartmut Neven, and Masoud Mohseni. Tensorflow quantum: a software framework for quantum machine learning. 2021. URL: <https://arxiv.org/abs/2003.02989>, arXiv:2003.02989.

[CT17]

Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017. URL: <https://arxiv.org/abs/1606.02318>.

[CBSG17]

Andrew W. Cross, Lev S. Bishop, John A. Smolin, and Jay M. Gambetta. Open quantum assembly language. 2017. URL: <https://arxiv.org/abs/1707.03429>, arXiv:1707.03429.

[DFO20]

Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020. <https://mml-book.github.io/book/mml-book.pdf>.

[DHM+18]

Danial Dervovic, Mark Herbster, Peter Mountney, Simone Severini, Naïri Usher, and Leonard Wossnig. Quantum linear systems algorithms: a primer. 2018. [arXiv:1802.08227](https://arxiv.org/abs/1802.08227).

[GBC16]

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.

[GK21]

Peter Groszkowski and Jens Koch. Scqubits: a python package for superconducting qubits. *arXiv e-prints*, jul 2021. [arXiv:2107.08552](https://arxiv.org/abs/2107.08552).

[HHL09]

Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, Oct 2009. URL: <https://arxiv.org/abs/0811.3171v3>, doi:[10.1103/physrevlett.103.150502](https://doi.org/10.1103/physrevlett.103.150502).

[HavlíčekCorcolesT+19]

Vojtěch Havlíček, Antonio D Corcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, Mar 2019. URL: <https://arxiv.org/abs/1804.11326>, doi:[10.1038/s41586-019-0980-2](https://doi.org/10.1038/s41586-019-0980-2).

[HSchatzleNoe20]

Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic schrödinger equation. *Nature Chemistry*, 12(10):891–897, 2020. URL: <https://www.nature.com/articles/s41557-020-0544-y>.

[HAGH+20]

Mohamed Hibat-Allah, Martin Ganahl, Lauren E Hayward, Roger G Melko, and Juan Carrasquilla. Recurrent neural network wave functions. *Physical Review Research*, 2(2):023358, 2020. URL: <https://arxiv.org/abs/2002.02973>.

[HZL+17]

He-Liang Huang, You-Wei Zhao, Tan Li, Feng-Guang Li, Yu-Tao Du, Xiang-Qun Fu, Shuo Zhang, Xiang Wang, and Wan-Su Bao. Homomorphic encryption experiments on ibm's cloud quantum computing platform. 2017. [arXiv:1612.02886](https://arxiv.org/abs/1612.02886).

[Isi25]

Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[KYG+07]

Jens Koch, Terri M Yu, Jay Gambetta, A A Houck, D I Schuster, J Majer, Alexandre Blais, M H Devoret, S M Girvin, and R J Schoelkopf. Charge-insensitive qubit design derived from the Cooper pair box. *Physical Review A*, 76(042319). 2007. [doi:10.1103/physreva.76.042319](https://doi.org/10.1103/physreva.76.042319).

[KKY+19]

P Krantz, M Kjaergaard, F Yan, T P Orlando, S Gustavsson, and W D Oliver. A Guide to Superconducting Qubits for New Quantum Information Engineers. *Applied Physics Review*, 2019. [arXiv:1904.06560v1](https://arxiv.org/abs/1904.06560v1).

[LLKG21]

Jing Li, Song Lin, Yu Kai, and Gongde Guo. Quantum k-nearest neighbor classification algorithm based on hamming distance. 2021. [arXiv:2103.04253](https://arxiv.org/abs/2103.04253).

[Lom04]

Chris Lomont. The hidden subgroup problem-review and open problems. *arXiv preprint quant-ph/0411037*, 2004.

[MBK21]

Andrea Mari, Thomas R. Bromley, and Nathan Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *Physical Review A*, 103(1):012405, Jan 2021. URL: <https://arxiv.org/abs/2008.06517>, [doi:10.1103/physreva.103.012405](https://doi.org/10.1103/physreva.103.012405).

[MST17]

Kyle Mills, Michael Spanner, and Isaac Tamblyn. Deep learning and the schrödinger equation. *Physical Review A*, 96(4):042113, 2017. URL: <https://arxiv.org/abs/1702.01361>.

[MKF19]

Kosuke Mitarai, Masahiro Kitagawa, and Keisuke Fujii. Quantum analog-digital conversion. *Phys. Rev. A*, 99:012301, Jan 2019. URL: <https://link.aps.org/doi/10.1103/PhysRevA.99.012301>, [doi:10.1103/PhysRevA.99.012301](https://doi.org/10.1103/PhysRevA.99.012301).

[MNKF18]

Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, Sep 2018. URL: <https://arxiv.org/abs/1803.00745>, [doi:10.1103/PhysRevA.98.032309](https://doi.org/10.1103/PhysRevA.98.032309).

[MRT18]

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2 edition, 2018. ISBN 978-0-262-03940-6.

[NSS+08]

Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, and Sankar Das Sarma. Non-abelian anyons and topological quantum computation. *Reviews of Modern Physics*, 80(3):1083–1159, Sep 2008. URL: <http://dx.doi.org/10.1103/RevModPhys.80.1083>, [doi:10.1103/revmodphys.80.1083](https://doi.org/10.1103/revmodphys.80.1083).

[RGM+17]

Marco Roth, Marc Ganzhorn, Nikolaj Moll, Stefan Filipp, Gian Salis, and Sebastian Schmidt. Analysis of a parametrically driven exchange-type gate and a two-photon excitation gate between superconducting qubits. *Physical Review A*, 2017. [doi:10.1103/PhysRevA.96.062323](https://doi.org/10.1103/PhysRevA.96.062323).

[SZK+18]

N. Samkharadze, G. Zheng, N. Kalhor, D. Brousse, A. Sammak, U. C. Mendes, A. Blais, G. Scappucci, and L. M. K. Vandersypen. Strong spin-photon coupling in silicon. *Science*, 359(6380):1123–1127, 2018. URL: <https://www.science.org/doi/abs/10.1126/science.aar4054>, [doi:10.1126/science.aar4054](https://doi.org/10.1126/science.aar4054).

[SCZ17]

Robert S. Smith, Michael J. Curtis, and William J. Zeng. A practical quantum instruction set architecture. 2017. [arXiv:1608.03355](https://arxiv.org/abs/1608.03355).

[SF14]

Leonard Susskind and Art Friedman. *Quantum mechanics: the theoretical minimum*. Basic Books, 2014.

[SYG+20]

Yudai Suzuki, Hiroshi Yano, Qi Gao, Shumpei Uno, Tomoki Tanaka, Manato Akiyama, and Naoki Yamamoto.

Analysis and synthesis of feature map for kernel-based quantum classifier. *Quantum Machine Intelligence*, 2(1):1–9, Jul 2020. URL: <https://arxiv.org/abs/1906.10467>, doi:10.1007/s42484-020-00020-y.

[VD17]

Uri Vool and Michel Devoret. Introduction to quantum electromagnetic circuits. *International Journal of Circuit Theory and Applications*, 45(7):897–934, 2017. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cta.2359>, doi:10.1002/cta.2359.

[WSB+05]

A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf.

Approaching unit visibility for control of a superconducting qubit with dispersive readout. *Physical Review Letters*, 95(6):060501, 8 2005. URL: <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.95.060501>, arXiv:0502645, doi:10.1103/PhysRevLett.95.060501.

[WBD+19]

K. Wright, K. M. Beck, S. Debnath, J. M. Amini, Y. Nam, N. Grzesiak, J. S. Chen, N. C. Pisenti, M. Chmielewski, C. Collins, K. M. Hudek, J. Mizrahi, J. D. Wong-Campos, S. Allen, J. Apisdorf, P. Solomon, M. Williams, A. M. Ducore, A. Blinov, S. M. Kreikemeier, V. Chaplin, M. Keesan, C. Monroe, and J. Kim. Benchmarking an 11-qubit quantum computer. *Nature Communications*, 10(1):1–6, dec 2019. arXiv:1903.08181, doi:10.1038/s41467-019-13534-2.

[01]

Кокин А. А. Валиев К. А. *Квантовые компьютеры: надежды и реальность*. НИЦ "Регулярная и хаотическая динамика", 2001.

[89]

Лифшиц Е. М. Ландау Л. Д. *Квантовая механика: Нерелятивистская теория*. Наука, 1989. URL: <https://www.math.purdue.edu/~eremenko/dvi/LL.pdf>.

[15a]

Иванов М.Г. *Как понимать квантовую механику*. Регулярная и хаотическая динамика, 2015. URL: <https://mipt.ru/upload/medialibrary/533/quant-2.pdf>.

[07]

Килин С. Я. Могилевцев Д. С. *Методы квантовой оптики структурированных резервуаров*. Белорусская наука, 2007. URL: <https://www.litres.ru/d-s-mogilevcev/metody-kvantovoy-optiki-strukturirovannyh-rezervuarov-7073607/>.

[15b]

Леонард Сасскинд and Арт Фридман. Квантовая механика. теоретический минимум. 2015.

Список авторов курса

Создатели курса

- [Семен Синченко](#)
- [Юрий Кашницкий](#)
- [Виктор Трохименко](#)

Основные авторы

- [Александр Березутский](#)
- [Александра Овсянникова](#)
- [Алексей Озерин](#)
- [Алексей Пронькин](#)
- [Андрей Лукьяненко](#)
- [Борис Зимка](#)
- [Вадим Петров](#)

- [Дани Эль-Айасс](#)
- [Дмитрий Бажанов](#)
- [Евгений Желтоножский](#)
- [Игорь Котенков](#)
- [Игорь Токарев](#)
- [Илья Беседин](#)
- [Карина Решетова](#)
- [Николай Карелин](#)
- [Никита Щуцкий](#)
- [Семен Синченко](#)
- [Сергей Ширкин](#)
- [Юрий Кашницкий](#)

Основные ревьюеры

- [Борис Зимка](#)
- [Виктор Трохименко](#)
- [Евгений Желтоножский](#)
- [Николай Карелин](#)
- [Роман Тезиков](#)
- [Юрий Кашницкий](#)
- [Ivan Dikhtyar](#)

Техническая поддержка

- [Алексей Пронькин](#)
- [Дмитрий Коржов](#)

Редакторы

- [Екатерина Пацкова](#)
- [Наталья Маркова](#)

Благодарности

Команда курса выражает благодарность:

- сообществу [Open Data Science](#) за предоставление платформы для размещения курса и техническую поддержку;

Глоссарий

А

Алгоритм Гровера

Англ. Grover's algorithm

Алгоритм квантового приближения для задачи оптимизации

(QAOA, quantum approximate optimization algorithm), <https://www.osp.ru/os/2019/03/13055118>,
<https://habr.com/ru/post/513474/>

Аннилер

Англ. annealer. Устройство для приближенного решения задачи оптимизации

https://en.wikipedia.org/wiki/Quantum_annealing

Б

Бра-вектор

Англ. bra-vector

B

Вакуумное состояние

Англ. Vacuum state, также используется название "основное состояние" (basic state). Так часто называется квантовое состояние с вектором $|\langle\text{ket}\{0\}\rangle|$. Термин заимствован из квантовой оптики, чаще всего не имеет ничего общего с "обыденным" понятием "вакуума".

Вариационная квантовая схема

Англ. Variational Quantum Circuits (VQC)

Вариационное машинное обучение

Англ. variational machine learning

Г

Гейты

Англ. Quantum gates

Гейт Адамара

Англ. Hadamard gate

Гейт C-NOT

Гейт NOT

T-гейт

Гейты поворота

Phase-shift гейт

Единичный гейт

Гейт iSWAP

fSim

И

Измерение

Англ. Quantum measurement

К

Квантовая информация

Также используется термин "квантовая информатика", англ. Quantum Information

Квантовый оператор

Англ. Quantum operator

Квантово-классическое обучение

Англ. Quantum-classical machine learning

Квантовое состояние

Англ. Quantum state

Кет-вектор

Англ. ket-vector

Кот Шредингера

Англ. Schrödinger's cat

Кубит

Также кубит или q-бит, англ. Qubit, от quantum bit. Минимальная единица информации на квантовом компьютере. Как и (классический) бит, кубит допускает два состояния, которые обычно обозначаются кет-векторами $|\langle\text{ket}\{0\}\rangle|$ и $|\langle\text{ket}\{1\}\rangle|$, а на уровне реализации кубит – это двухуровневая квантовая система, такая как две поляризации фотона или спина электрона.

Кудит

Англ. qudit. Квантовая система с d состояниями.

M

Матрица плотности

Также используется термин "оператор плотности", англ. Density matrix или Density operator

Матричное прямое произведение

Синоним произведение Кронекера см. Кронекер Англ. Tensor product

Метод квантового локального подбора

Многокубитные гейты

Англ. Multiqubit gates

Н

Нотация Дирака

Англ. Dirac notation, формализм (система обозначения) для описания квантовых состояний на основе бра- и кет-векторов (англ. bra-ket от bracket, скобка). В этой системе обозначений $\langle \Psi |$ обозначает вектор состояния, $\langle \Psi |$ – сопряженный вектор состояния, а $\langle \Psi | \hat{H} | \Psi \rangle$ – среднее (математическое ожидаемое) значения наблюдаемой для оператора \hat{H} в состоянии $|\Psi\rangle$ (также говорят о свертке оператора $\langle \hat{H} |$ с бра-вектором $\langle \Psi |$ и кет-вектором $| \Psi \rangle$). Система обозначений Дирака позволяет отделить отдельно вектора состояния $|\Psi\rangle$ от самого состояния $\langle \Psi |$ или конкретного математического описания (в виде волновой функции или вектора).

О

Операторы Паули

Также используется термин “матрицы Паули”, англ. Pauli operators или Pauli matrices

П

Правило сдвига параметров

(*) Англ. Parameter-shift rule

Произведение Кронекера

тоже, что и матричное прямое произведение (тензорное произведение в случае линейных операторов) (*)

Англ. Kronecker product

С

Смешанные состояния

Англ. Mixed states

Собственный вектор

Англ. Eigenvector

Собственное значение

Англ. Eigenvalue

Соотношение неопределенности

Также используется термин “принцип неопределенности”, англ. Uncertainty relation или Uncertainty principle

Суперпозиция

Англ. Superposition

Сфера Блоха

Англ. Bloch sphere

Т

У

Унитарный оператор

Англ. Unitary operator

Ш

«Шумные» квантовые компьютеры промежуточного масштаба (Noisy intermediate-scale quantum, NISQ, NISQ Hardware)

Термин введен Прескиллом в работе <https://arxiv.org/abs/1801.00862> для описания сегодняшнего состояния квантовых вычислений, когда “настоящие” квантовые вычисления (на миллионах кубит) еще невозможны, но современные квантовые компьютеры уже могут выполнять вычисления, которые невозможны на обычных,

классических компьютерах. См. также краткое обсуждение в
<https://quantumcomputing.stackexchange.com/questions/1885/what-is-meant-by-noisy-intermediate-scale-quantum-nisq-technology>, и перевод на термины русский <https://www.osp.ru/os/2019/03/13055130>

Э

Энтропия

Англ. Entropy

By ODS Quantum Community

© Copyright 2021.