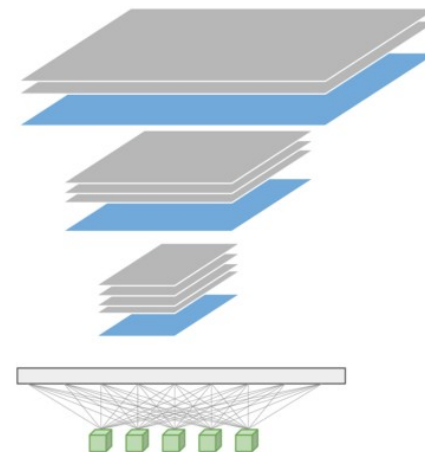
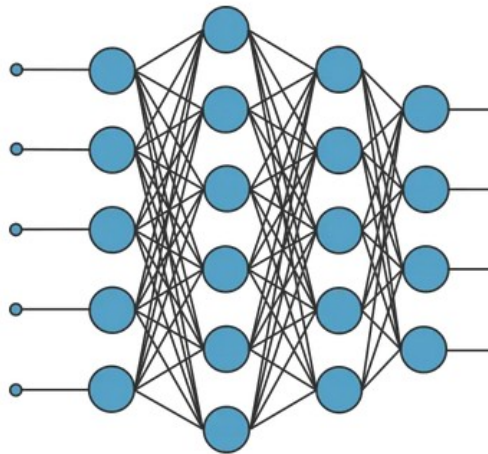
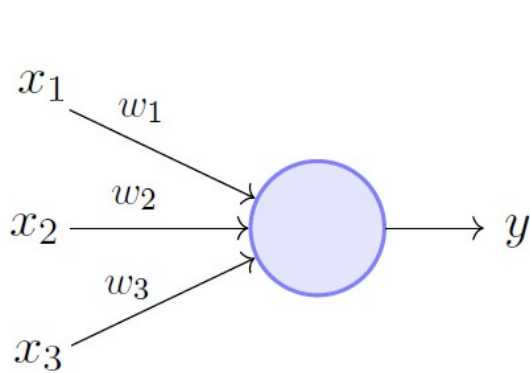


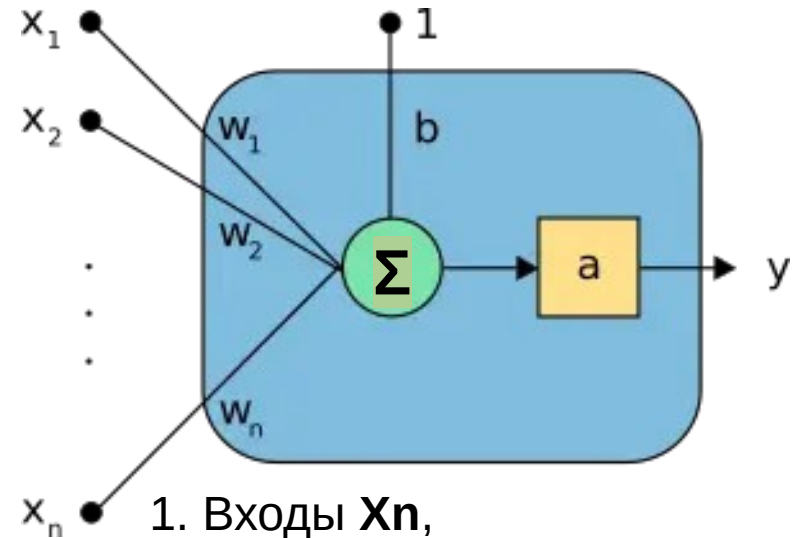
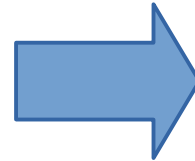
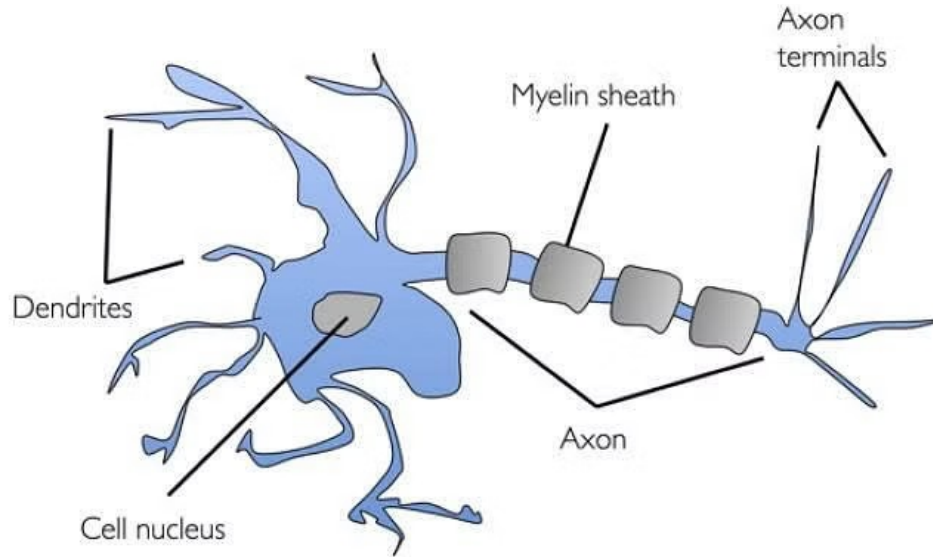
Нейронные сети: от математических моделей к практическим методам

(Краткая история ИИ)



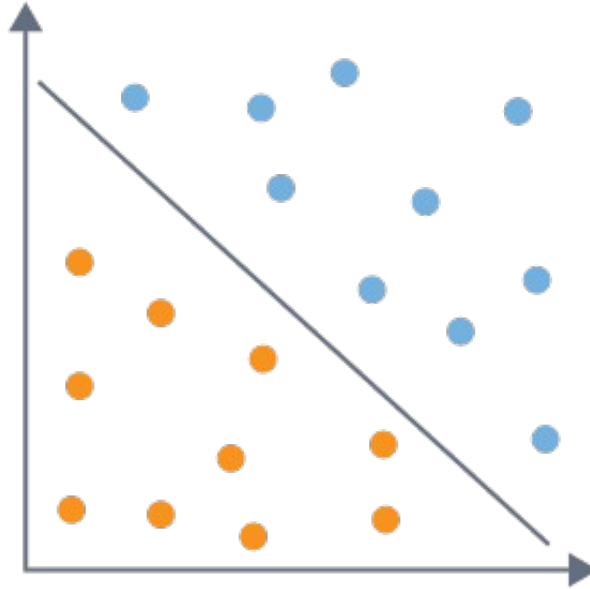
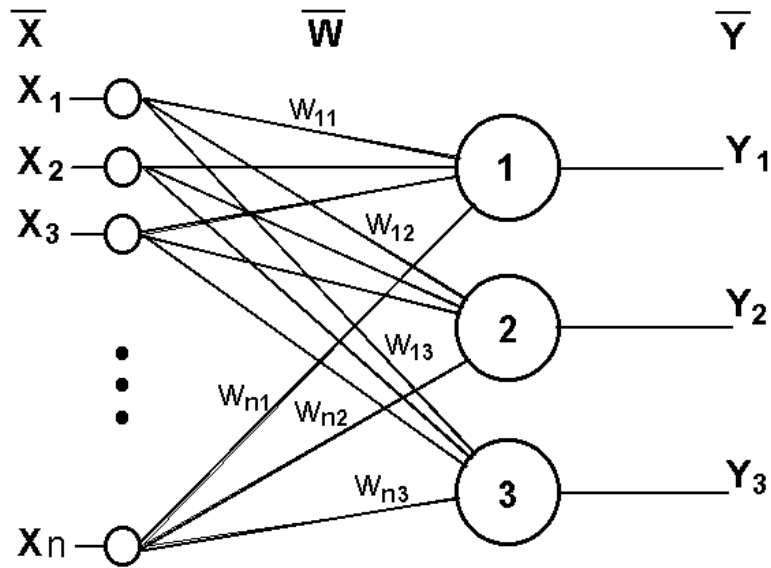
1950-е: теория нейрона

- Одна из первых моделей нейрона (МакКаллок, Питтс, 1943)
- Работа на алгоритме “пороговой логики”



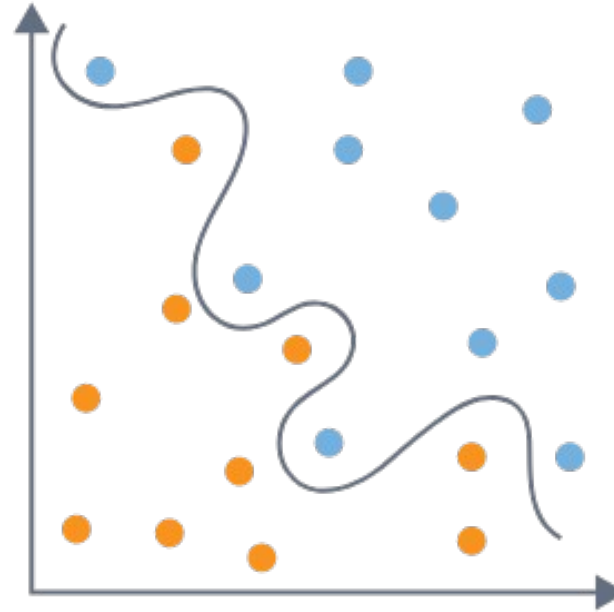
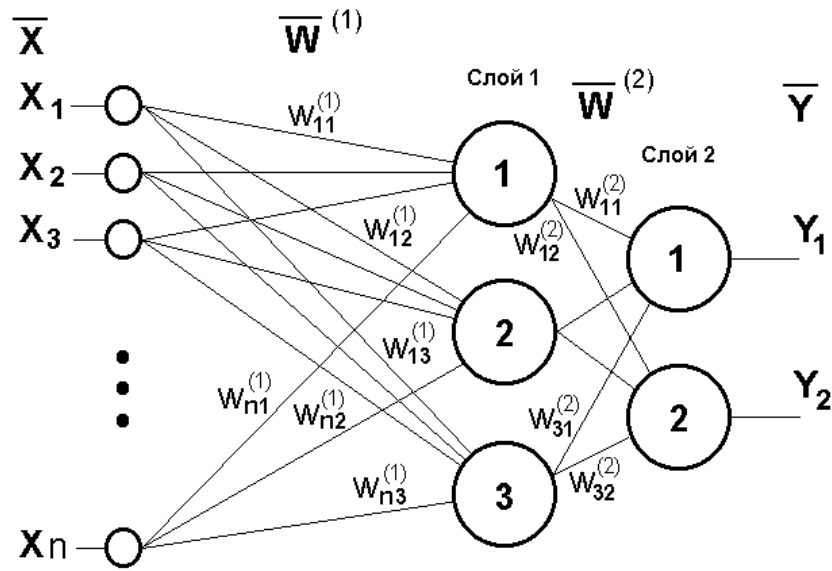
- Искусственный нейрон – математическая модель биологического нейрона
- Элементарный вычислительный узел
- **Возможности:** бинарная классификация числовых значений

1960-е: однослойный персептрон



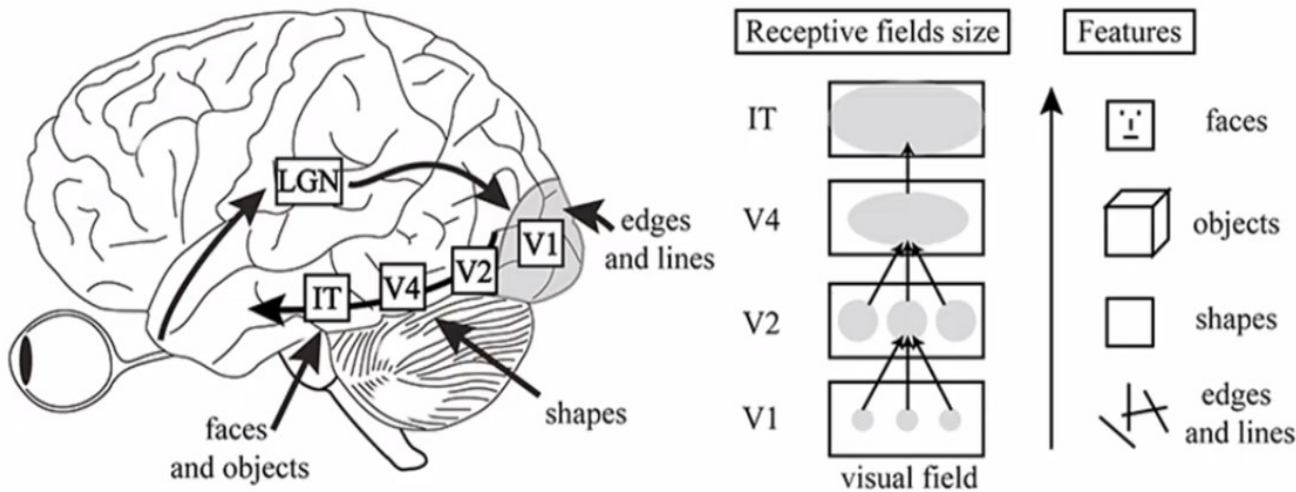
- Однослойный персептрон (SLP – single layer perceptron, Розенблатт, 1957). Реализован в компьютерной модели “Марк-1”, 1960)
- Полносвязность: промежуточные нейроны связаны с каждым предыдущим
- “Обучение” на примерах: алгоритм обратного распространения ошибки (backprop)
- **Возможности:** бинарная классификация линейно разделимых числовых векторов

• 70-е: многослойный персептрон



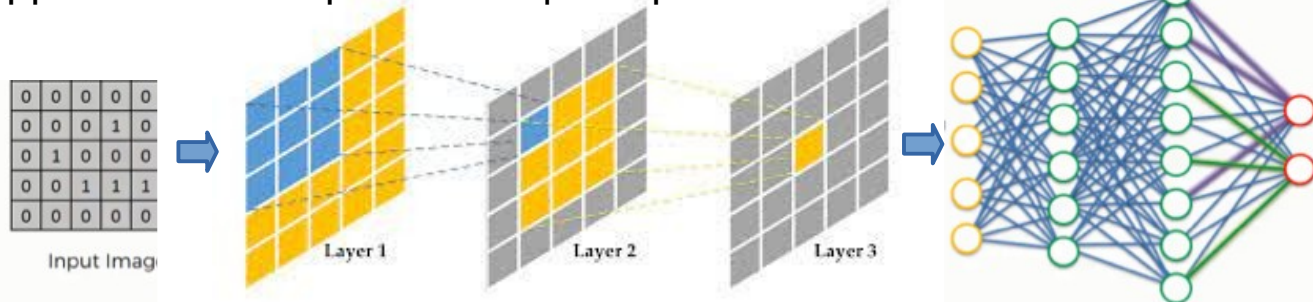
- Кроме входного один или более промежуточных слоев
- Полносвязность: промежуточные нейроны связаны с каждым предыдущим
- “Обучение”: алгоритм обратного распространения ошибки
- **Возможности:** классификация линейно неразделимых числовых векторов

Конец 70-х: сверточные нейронные сети



Из Manassi, Sayim, Herzog, 2013

Два слоя со сверточными фильтрами 3x3:



- **Возможности:** классификация изображений

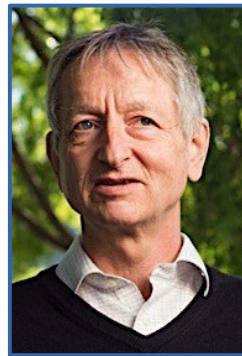
- Сверточные сети (Convolutional neural networks, CNNs) – архитектура сети, имитирующая анализ зрительной информации мозгом млекопитающих
- Иерархия сверточных фильтров (аналог рецептивных полей) и обобщение признаковой информации при прохождении по слоям сети
- Неполносвязная сеть: к следующему слою идет небольшое число связей от предыдущего

Конец 70-х - 90-е: “Зима ИИ”

- Персептрон не способен работать со сложными данными (Minsky, Papert, 1969)
- Завершение эпохи “больших экспертных систем”, снижение оптимизма от ИИ
- Проблема “обнуления / взрыва градиентов” при обучении многослойных НС
- НС работают, но обучать их очень долго и дорого по вычислительным ресурсам

...it would take more than all of the world's computers the rest of humanity's existence to train a single neural network!

G. Hinton, 1986

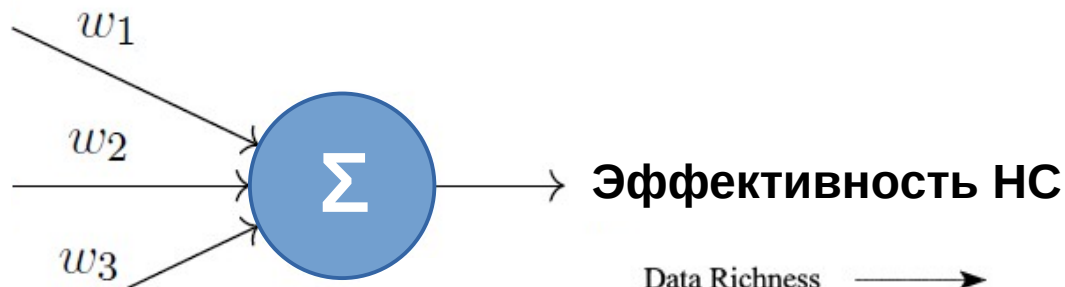


1990-2000: “накопление потенциала”

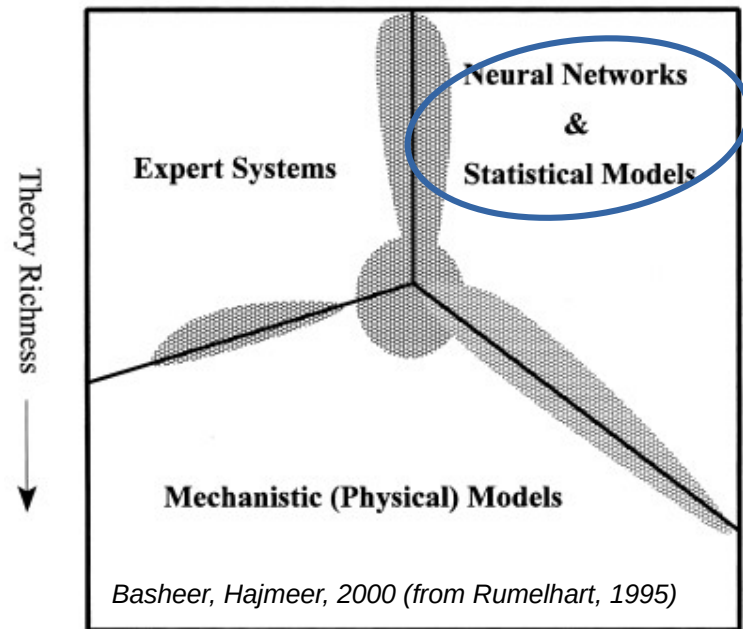
Алгоритмы обучения, архитектуры сетей. **Преодоление проблемы вычисления градиентов**

Вычислительные мощности.
“Закон Мура” (1986: 10^5 , 2006: 10^{12} транзисторов)

Большие объемы данных: развитие интернета, мобильных устройств.
“Мир данных”



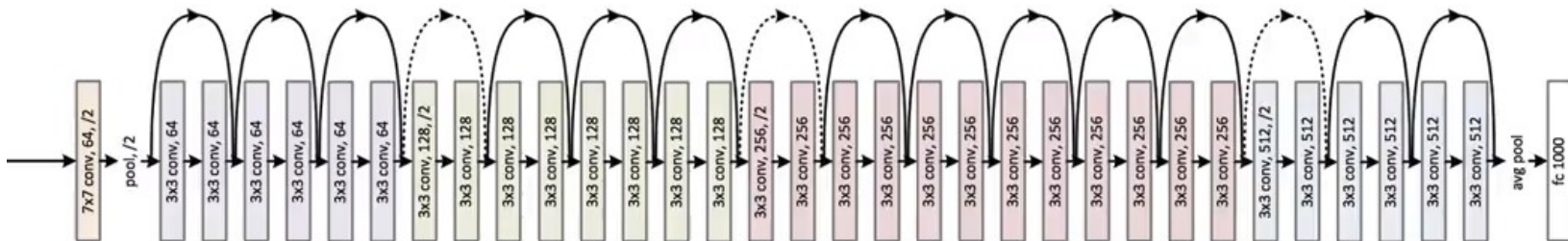
Data Richness →



2010-е: “Весна ИИ”

Триумф CNNs в задачах компьютерного зрения:

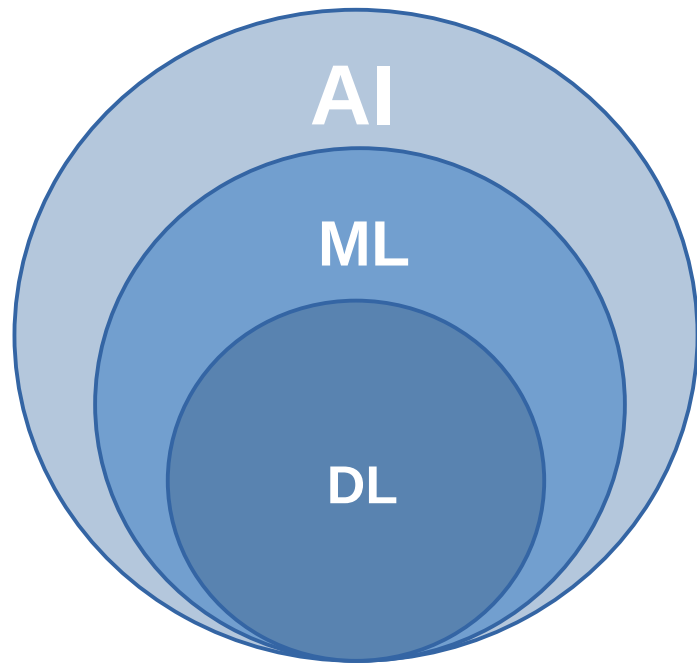
- Архитектура LeNet (LeCun et al., 2008)
- Создан набор данных ImageNet (2009): размечено 14 млн изображений, 1000 классов (Li Fei Fei et al., 2021)
- Архитектура AlexNet (Krizhevsky et al., 2012): эффективное решение проблемы расчета градиентов при обучении глубоких сетей, расчеты на видеопроцессорах (GPU)
- Семейство сетей архитектуры ResNet (He et al., MS Research, 2015), обошла по точности человека на ImageNet
- Архитектуры YOLO (Redmon et al., 2015), UNet (Ronnenberger et al., 2015),



Архитектура
ResNet-34

Сейчас: “Лето ИИ”

- **Искусственный интеллект** (Artificial intelligence, AI) – область разработки и исследования методов по аналогии с человеческими решениями, широкая междисциплинарная сфера, нет четких границ
- **Машинное обучение** (Machine learning, ML) – область применения математических методов и “обучаемых” алгоритмов для решения прикладных задач
- **Глубокое обучение** (Deep learning, DL) – область применения математических методов и “обучаемых” алгоритмов на основе многослойных НС.
- **Триумф “коннекционизма”**: искусственные НС помогут понять мышление и сознание



Глубокое
обучение

→ “Обучение с учителем” (*Supervised learning*)

= обучение на примерах, нужна разметка “правильных ответов” на исходных данных

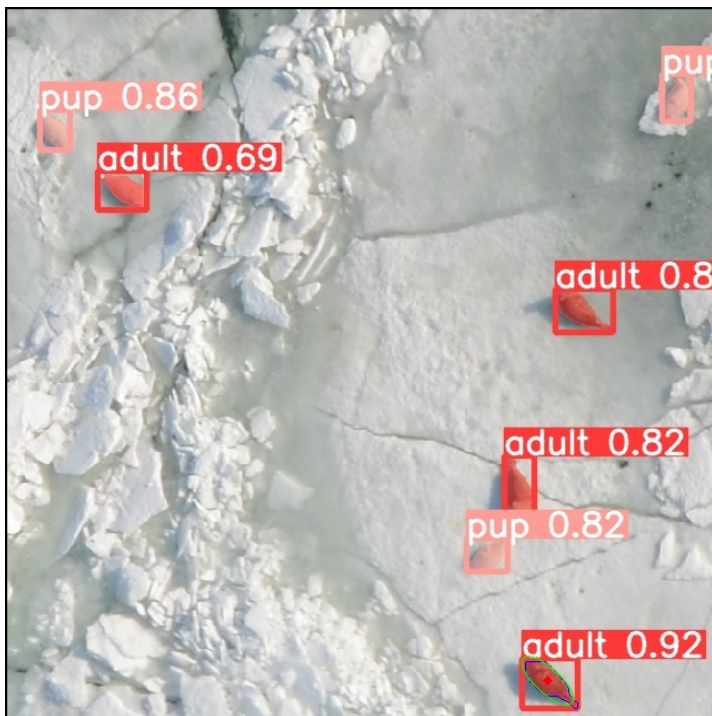
→ “Обучение без учителя” (*Unsupervised learning*)

Расширение класса задач

- Классификация изображений (VGG, ResNet)
- Обнаружение объектов (Faster RCNN, SSD, YOLO)
- Сегментация объектов + классификация (UNet, DeepLab)
- Работа с видео, в т.ч. в реальном времени (YOLO)

Компьютерное зрение

Пример работы обученной модели архитектуры YOLOv8s для задачи сегментации объектов (авиаучет тюленей)

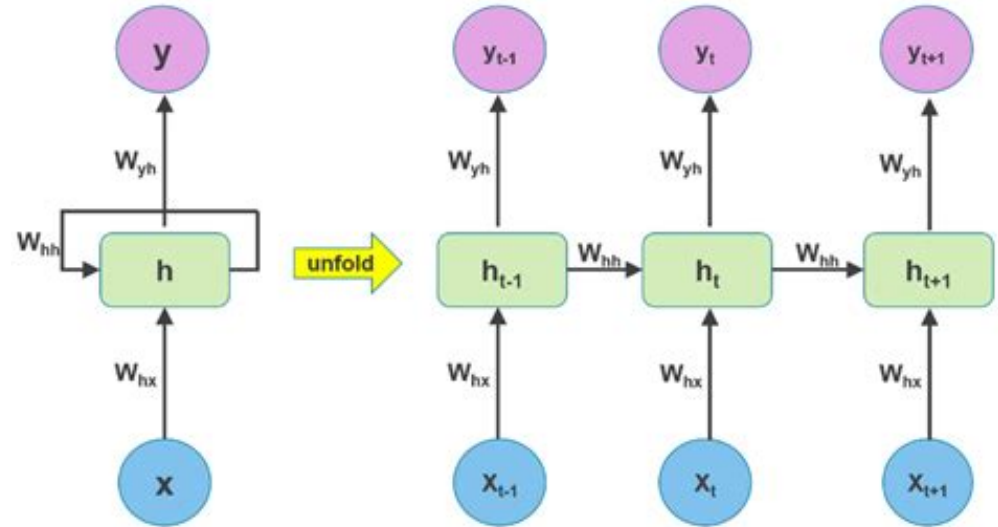


Расширение класса задач

Обработка символьных последовательностей, временных рядов

Реккурентные сети (RNNs) и Long Short Term Memory (LSTM, Hochreiter, Schmidhuber, 1997)

- Некоторые узлы (“нейроны”) имеют петли обратной связи
- Обратная связь – основа механизма внутренней “памяти” состояния



- **Возможности (задачи):**

- Text-to-text: машинный перевод
- “Сочинение”, дополнение фраз, аннотации к тексту (GPT-3,4, LLMs)
- “Диалог” (ChatGPT)
- Обработка звука, ЭКГ, ЭЭГ и других временных сигналов
- Поиск, предсказание структур биомолекул (AlphaFold)

Обучение с подкреплением (Reinforcement Learning) – парадигма машинного обучения, в которой множество агентов учатся совершать действия в эмулируемой среде по определенным правилам с помощью системы “наград” и “штрафов”

- 1997: Deep Blue (еще “классический” программируемый алгоритм) побеждает Каспарова в шахматы
- 2016: AlphaGo побеждает Ли Седоля (чемпиона мира в Го).
- 2017: AlphaZero после 2 часов самообучения играет в шахматы и го лучше любого из людей, после 4 часов – лучше любой из программ (в т.ч. AlphaGo, обучающейся 3 дня)
- 2020: MuZero успешно обучается играть в видеоигры на приставках, а также в настольные игры, не зная правил заранее
- 2020: AlphaStar: победа лучших игроков в компьютерной стратегии реального времени StarCraft 2

Расширение класса задач

Мультимодальные архитектуры

- Генеративные модели для частных задач
- Text-to-image / video: генерация изображений и видео по текстовому описанию (GANs, Diffusion models)
- time series-to-text: предсказание содержания внутренней речи по записи многоканальной ЭЭГ (hybrid / transformers)
- time series-to-image: предсказание содержания рассматриваемого объекта по записи многоканальной ЭЭГ (hybrid / transformers)
- Диалог с поддержкой мультимодальности (ChatGPT-4)

*Генеративная сеть StyleGAN 2:
генерация лиц несуществующих людей*

<https://thispersondoesnotexist.com/>



Примеры (неудач)

Попытки сообщества исследовать Covid-19 с 2020 г:

- НС для раннего выявления болезни, предсказания выздоровления / осложнений, масса моделей предсказания роста заболеваемости в разных странах...
- Множество моделей для работы с данными различной модальности: тексты, сиквенсы, изображения, цифровые данные
- Исследование *The British Medical Journal* и *Nature Machine Intelligence*: на середину 2021 г **ни одна** из моделей на основе ИИ не помогает в борьбе с Covid-19 и не может использоваться в клинических условиях!

Причины неудач (и проблем ИИ в целом)

- **Garbage in – garbage out.** Главная проблема – данные для обучения: дубликаты, артефакты, смещенность выборок, неточность ручной разметки
- **“Волшебный черный ящик”.** Непонятно как модель принимает решения (общая проблема глубокого обучения). Современные модели уже крайне сложны для детального реверс-анализа, сложность постоянно растет
- **Человеческий фактор:** желание выделиться (“модные” методы, глобальная проблема...). Вместо объединения усилий для решения проблемы – конкуренция за ресурсы и внимание

Новый вид угроз из-за умышленных действий: deep fake, adversarial data injection и другие виды атак на работу DL-алгоритмов

Текущее состояние методов глубокого обучения

- Масса готовых (и открытых) архитектур по многим типовым задачам из разных областей, не нужно разрабатывать с нуля
- Метод переноса обучения (transfer learning), можно сохранить и поделиться файлом состояния обученной модели, с этим не нужно обучать с нуля, только “доучить”
- Богатый программный инструментарий открытого доступа для работы с НС
- Большое число наборов обучающих данных (в т.ч. открытых) в различных областях
- Доступность вычислительных мощностей (эффективное распараллеливание на графических ускорителях)

Типы данных и типы задач

- числовые,
- символьные (тексты естественного языка, сиквенсы биомолекул),
- матричные / тензорные (изображения, в т.ч. мультиспектральные),
- временные ряды (звук, ЭКГ, ЭЭГ и др), временной ряд матриц (видео)

Модальность	Тип сигнала	Тип данных	НС-архитектура
Зрение	Изображение Видеоряд	Матрица Временной ряд матриц	CNNs
Слух	Звук	Временной ряд	RNNs / LSTM
Язык*	текст, речь	Символьная последовательность	RNNs / LSTM
Мультимодальные	Текст, звук, изображение...	Разные	Hybrid / Transformers

- Классификация элементов данных
- Обнаружение / сегментация: выделение части элемента для всех элементов
- Выделение общих признаков всех элементов
- Генерация синтетических элементов / предсказание теоретически возможных
- Соотнесение элементов одного типа данных с элементами другого

Тенденции развития глубокого обучения

НС как универсальные “ассоциативные машины”: универсализация по типам данных и дальнейшее развитие больших мультимодальных НС (text-to-image, sound-to-image и т.д.). Создание эффективных архитектур и алгоритмов для “сложных” модальностей

Data-centric programming. Масштабные коллекции обученных “микро-моделей” под узкие частные задачи. В новой парадигме программирования микро-модели в статусе объектов классической парадигмы ООП

Unsupervised (semi-supervised) learning. Развитие методов “обучения без учителя” (без явной разметки примеров правильных ответов или с минимальной разметкой)

Interpretable AI. Разработка методов реверс-анализа принципов принятия решений сложными моделями. Из “черного ящика” хотя бы “серый”.

Tiny ML (Edge AI). Развитие аппаратных средств и платформ разработки обучаемых моделей на мобильных / носимых устройствах, автономных датчиках. Сети “умных” датчиков.

Предложения и рассуждения по AGI (Artificial General intelligence, “Сильный ИИ”). Иерархическая концепция, концепция нейро-подобия, акцент на обучении без учителя и другие.

Вместо выводов

- То что создавалось как модель работы мозга, по прошествии 60 лет развития стало мощным и уже общепринятым инструментом анализа научных данных...
- Принцип верификации по оригиналу: если модель работы мозга существенно облегчает работу оригинала – модель не просто верна, она успешна в практическом плане
- Методы ИИ это в первую очередь инструменты. Как молоток – специализированное “продолжение” руки, так и эти методы – “продолжение” когнитивных способностей человека. Но это универсальный и очень мощный инструмент...

“I often tell my students not to be misled by the name ‘AI’ – there is nothing artificial about it. AI is made by humans, intended to behave by humans, and, ultimately, to impact humans’ lives and human society.”

Fei Fei Li



Благодарю за внимание!

All models are wrong... but some are useful

G. Box, 1976

