

Организационные вопросы

Семинар 1

Источники информации

- [Страница курса](#)
- [YouTube-канал с лекциями](#)
- [Канал в Telegram](#)
- [Open Data Science \(ODS\) slack](#)
- Канал курса в ODS slack: **#dlcourse_ai**
- Канал курса в CSC slack: **#dl-simple-19**
- Канал на [ClosedCircles.com](#)

Общая информация

- К семинаристам обращаться на **ТЫ**
- **Обязательно** смотреть лекции - они классные
- Семинары CSC **НЕ** записываются на видео
 - Но записываются семинары одной из групп НГУ, следите за каналом курса в ODS
- У нас жесткие дедлайны
- Последние 1 или 2 семинара будут длинной в 2 пары

Необходимые инструменты

- Python 3.6+
- Jupyter Notebook (Google Colab)
- Base libraries
 - [numpy](#), [matplotlib](#), [scikit-learn](#)
- Deep learning
 - CUDA, cuDNN
 - [pytorch](#)
- Рекомендуем озаботиться подготовкой окружения заранее
- Рекомендуем пользоваться [venv](#)
 - как вариант - установить [Anaconda](#)

Какие вычислительные мощности нужны?

- Без GPU будет грустно
 - Но можно попытаться затрайхардить
- Рекомендуется
 - Свой компьютер с GPU от NVIDIA (хотя бы GTX 9**)
 - [Google Colab](#) ([гайд по настройке](#)). Теперь и с pytorch из коробки
- Другие варианты
 - Облачные сервисы: [AWS](#), [GCP](#), ...
 - Университетский кластер
 - Сервера на работе
 - Попросить доступ у друга
 - [Kaggle Kernels](#)

Нужен ли ноутбук на семинаре?

Желателен, но необязателен

- Будут семинары, где будем писать код вместе
- Возможно, вам проще будет смотреть материалы у себя

Структура оценок

Система оценивания

Способы получить баллы:

- 6 домашних заданий: каждое от 0 до 8 баллов (максимум 48)
- Контрольная работа: от 0 до 10 баллов
- Разбор DL статьи: от 0 до 7 баллов
- Проект: от 0 до 35 баллов
- Пятиминутки: до 0.5 балла за каждую

В сумме получается максимум 100 баллов + пятиминутки (до 5-6 баллов)

Критерии оценок:

- Отлично: 80 баллов
- Хорошо: 60 баллов
- Зачет: 40 баллов
- Для СПб: -5 баллов на все

Домашние задания

- Максимум 8 баллов
- Python 3
- Сдавать на сайте CSC, если не работает - в личку slack
- Как получить **0 баллов** за полностью выполненное задание?
 - Просрочить дедлайн на 1нс
 - Удалить условия заданий
 - Плагиат
- Пожелания к оформлению
 - Фамилия и имя в названии файла
 - Если начиная с задания N ничего не делалось - проверяющий будет благодарен, если вы напишите об этом **ОЧЕНЬ ЗАМЕТНО**
 - Если не делалось задание, то **АНАЛОГИЧНО**
 - У проверяющего проблемы с различением зеленого, красного, коричневого и серого цветов

Домашние задания: система штрафов

Кроме очевидных ошибок, вроде неверно выполненных заданий, баллы снимаются за следующее:

- Нет разумных и понятных выводов, где они требуются
- Совсем уж нечитаемый код
- Неиспользуемый код, много закомментированного кода
- Warnings, errors
- Непонятные графики, графики без пояснений, пояснения без графиков
- Совмещение условий разных заданий в одно, перестановка заданий местами
- Черисмерные ашипки, в приминение, рускава изыка (end Inglesch tu)

Домашние задания: темы

- **Assignment 1 (семинар 2)**
 - KNN, implement metrics: precision, recall, F1
 - Linear Classifier, SGD, derive gradients
- **Assignment 2 (семинар 3)**
 - Neural nets, layers architecture, implement backprop
 - Same with PyTorch
- **Assignment 3 (семинар 6)**
 - Conv & maxpool layers and their backward pass.
 - CNNs with PyTorch
- **Assignment 4 (семинар 7)**
 - Transfer learning
- **Assignment 5 (семинар 9)**
 - Word2Vec
- **Assignment 6 (семинар 10)**
 - LSTM

Контрольная работа

- Максимум 10 баллов
- Проходит в середине семестра (**семинар 8, 15.04.2019**)
- Теоретические вопросы по базовым темам курса
- Во время контрольной работы можно пользоваться конспектами, интернетом и материалами лекций
- Ожидается, что каждый ответ сопровождается объяснением и выводом результата
- Можно написать заочно
 - **12:00 20.04.2019 - 12:00 21.04.2019** (предварительная дата, время московское)
 - Получить можно только $\frac{2}{3}$ баллов

Разбор DL статьи

- Максимум 7 баллов
- Требуется прочитать современную научную работу (обычно paper на arxiv) и коротко о ней рассказать на семинаре
 - Предполагаемое время выступления - около 10 минут
- Тему нужно согласовать с семинаристами **до семинара 9 (22.04.2019)**
 - Чем раньше выберете тему, тем лучше
 - Выступить нужно в течение семестра, не более двух человек за занятие
 - Нужно предупредить о выступлении хотя бы за неделю

Рекомендуемый план презентации:

1. Суть проблемы
2. Объяснить метод, предложенный в статье
3. Результаты и их интерпретация
4. Ответы на вопросы

Разбор DL статьи: примеры тем

- [Self-Normalizing Neural Networks](#)
 - Про эффективное обучение полносвязных нейронных сетей без регуляризации
- [How to Escape Saddle Points Efficiently](#)
 - Про время сходимости нейронных сетей
- [Flexpoint: An Adaptive Numerical Format for Efficient Training of Deep Neural Networks](#)
 - Про использование float16 для тренировки
- [On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima](#)
 - Про оптимальный размер минибатча
- [Training Deep Nets with Sublinear Memory Cost](#)
 - Про более эффективное хранение нейросетей в памяти
- [SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size](#)
 - Про сжатие нейросетей
- [Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions](#)
 - Про генерацию речи
- [Unsupervised Image-to-Image Translation Networks \(explaining deepfakes algorithm\)](#)
 - Про перенос параметров одного изображения на другое для фото и видео
- [A Neural Algorithm of Artistic Style](#)
 - Перенос стиля изображения
- [On the Dimensionality of Word Embedding](#)
 - Подбор размера word embeddings
- [How Does Batch Normalization Help Optimization?](#)
 - Статья с NeurIPS 2018 на тему “почему на самом деле работает BatchNorm”
- [Pelee: A Real-Time Object Detection System on Mobile Devices](#)
 - Новая быстрая и точная архитектура для Object Detection

Курсовой проект

- Максимум 35 баллов
- Самодостаточный практический проект в области deep learning, который выполняется одним человеком
 - в качестве исключения - группой из двух человек, но проект должен быть большего объема
- Проект нужно согласовать с семинаристами **до семинара 9 (22.04.2019)**
- Проект нужно будет презентовать на последнем семинаре **(20.05.2019)**

Что нужно продумать до начала проекта:

1. Цель работы
2. Сколько вашего и компьютерного времени займёт
3. Baseline-алгоритм
4. Какой датасет используется (если нет готового — то сколько времени займет его получить?)
5. Как решается ваша задача в различных научных (и не только) работах?

Курсовой проект: оценивание

- Реализован ли baseline?
- Реализованы ли какие-то улучшения baseline?
- Количество проверенных идей
- Качество и performance итогового решения
- Есть ли демо?
- Есть ли сравнение с другими решениями?
- Качество объяснения и презентации
- Качество ответов на вопросы

Курсовой проект: примеры

- Участие в текущем ML competition в области Deep Learning (например, на [Kaggle](#))
- Autocomplete для коротких сообщений а-ля [Google Smart Reply](#)
- RL для простой игры на реакцию (например, арканонид)
- Анализ тональности отзывов на кино или отзывов продуктов
- Запуск любой сети для классификации, сегментации или object detection на телефоне
- Классификация простых голосовых команд
- Простенький чат-бот
- Детекция светофоров на видео с видеорегистратора
- Text To Speech с конкретным голосом
- Deep fake с конкретным лицом (цензурный)
- ...

Курсовой проект: competition

Подробнее про ML competitions и Kaggle:

- Семинар 3 курса Machine Learning
- [Выступление Павла Плесскова на DataFest Siberia](#)
- [Пост в блоге Яндекса на хабре](#)
- [Пост в блоге ODS на хабре](#)
- [Канал ML-тренировок на YouTube](#)
- ODS: каналы **#kaggle_crackers**, **#mltrainings_beginners**, **#mltrainings_live**

Курсовой проект: competition

- По результатам соревнования нужно предоставить код и выступить с разбором своего решения
- Примеры выступлений можно посмотреть на [канале ML тренировок](#)
- Будет оцениваться не итоговое место, а ваш путь к получившемуся решению
 - Если медаль получена копипастой одного public kernel, то на много баллов рассчитывать не стоит
 - Но в то же время, грамотное использование public kernels, информации в ODS и других публичных источников - путь к успеху

Курсовой проект: competition

- Команды из одного человека
 - Исключение - реальная борьба за золото или деньги, но перед объединением нужно согласовать его с семинаристами
- Выбор соревнования нужно согласовывать
 - Лектор и семинаристы будут что-то рекомендовать
 - Но можно предложить и что-то свое
- Можно выбирать соревнования, которые заканчиваются после дедлайна
 - В этом случае презентуется разбор текущего решения

Пятиминутки

- Маленькие несложные задания, на которые нужно ответить письменно примерно за 5 минут
- Проводятся в начале семинара, пока ждем опоздавших
- Правила оценивания:
 - Каждая пятиминутка оценивается в максимум **0.5 балла**
 - За пропуск занятия или не сданную пятиминутку штрафов нет
 - Досдавать нельзя
 - Апелляции не принимаются
 - Опоздал - не повезло
- Стоит относиться, как к варианту самопроверки
- Стоит относиться, как к замене досдачам
- Не стоит относиться, как к источнику баллов
 - Статистика показывает, что вряд ли получится набрать больше 3 баллов

Дедлайны

- Домашние задания
 - 2 недели от даты выдачи
- Контрольная
 - Пишется очно на семинаре: **семинар 8, 15.04.2019**
 - Можно написать заочно вместе с СПб, но получите не более ½ баллов
- Разбор DL статьи
 - На семинарах в течение семестра, не более 2 за семинар, предупреждать минимум за неделю
 - Дедлайн согласования тем: **семинар 9, 22.04.2019**
 - Последний семинар с выступлениями: **семинар 12, 13.05.2019**
 - Но лучше выбрать (и рассказать) заранее
- Проект
 - Согласование тем: чем раньше, тем лучше вам
 - Дедлайн согласования тем: **семинар 9, 22.04.2019**
 - Презентация проекта: **семинар 13, 20.05.2019**

Санкт-Петербург

Для СПб

- Домашние задания
 - одинаковые правила и дедлайны с НСК
- Контрольная
 - Пишется заочно: **12:00 20.04.2019 - 12:00 21.04.2019** (предварительная дата, время московское)
 - Сдача в первые 2 часа - полный балл
 - Сдача после 2 часов от начала - $\frac{2}{3}$ баллов

Для СПб: разбор DL статьи

- Вместо выступления нужно сделать письменный разбор (примеры в канале ODS **#article_essense**)
- Дедлайн согласования тем: **22.04.2019**
- Дедлайн выполнения задания: **13.05.2019**
- Ваши разборы статей будут опубликованы в материалах курса

Для СПб: проект

- Только ML competitions
 - Другие виды проектов только в виде исключений и по предварительному согласованию
- Вместо выступления нужно предоставить код и письменный отчет
 - Примеры отчетов: [1](#), [2](#), [3](#), [4](#), [5](#) и много других в Discussions на Kaggle
- В отчетах нужно отразить следующее:
 - Обработка данных
 - Используемые модели, алгоритмы, трюки
 - Какие идеи заработали, а какие нет
 - Идеи, на которые не хватило времени
- Дедлайн согласования: **22.04.2019**
- Дедлайн предоставления отчетов: **20.05.2019**