



Семестр 2.

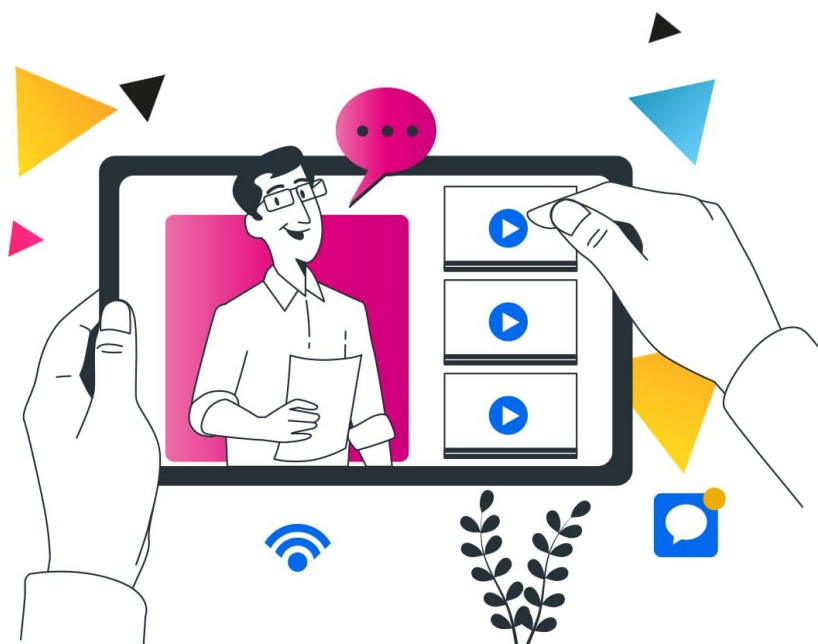
Машинное обучение

Машинное обучение и анализ больших данных



Содержание

Содержание	2
Общая информация о семестре	3
Преподаватели	3
О семестре	4
Цель курса	4
Образовательные результаты	4
Содержание курса	5
Тематический план	5
Форматы обучения	6
Каникулы	6
Оценивание	6
Формула оценки	6
Дополнительные сложные задачи	7
Дедлайны	8
Критерии успешного прохождения курса	11
Хакатон	12
Примеры расчета оценки	12



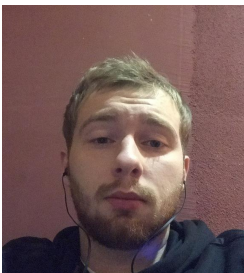




Общая информация о семестре

Название	Машинное обучение и анализ больших данных
Длительность	6 месяцев, 16 января - 30 июня
Семестр	2-й
Зачетные единицы	3
Аттестация	зачет

Преподаватели

	<p>Радослав Нейчев</p> <p>Читает лекции</p> <p>Старший преподаватель, Кафедра машинного обучения и цифровой гуманитаристики МФТИ</p>
	<p>Лавприт Сингх-Пальчевская</p> <p>Ведет семинары</p> <p>Младший научный сотрудник МГУ им. Ломоносова, кафедра биоинженерии</p>
	<p>Артём Островский</p> <p>Ведет семинары</p> <p>QKD researcher, CV engineer</p>



0 семестре

Второй семестр научит вас работать с большими массивами данных с помощью моделей машинного обучения. В семестр входят второй, третий и четвертый модули всей программы.

Второй модуль посвящен основам машинного обучения, а также математическому аппарату и инструментам, которые понадобятся для работы с данными. Третий модуль познакомит вас с прикладными моделями машинного обучения. В четвертом модуле вы рассмотрите модели глубокого обучения и закрепите навыки решения прикладных задач.

В результате освоения курса вы сможете решать реальные задачи с помощью методов машинного и глубокого обучения.

Цель курса

Научиться решать прикладные задачи с помощью методов машинного и глубокого обучения.

Образовательные результаты

В результате обучения на курсе вы научитесь:

- использовать математический аппарат для работы с данными;
- использовать Python для загрузки и предобработки данных;
- реализовывать базовые модели машинного обучения;
- использовать линейные модели и строить решающие деревья для решения задач анализа данных;
- оценивать качество построенной модели;
- оценивать значимость признаков;
- бороться с переобучением построенных моделей;
- работать с естественным языком и изображением с помощью нейросетевых моделей;
- использовать методы регуляризации нейронных сетей и аугментации данных;
- бороться с переобучением нейросетей.



Содержание курса

Тематический план

№	Тема
Модуль 2. Математические основы машинного обучения	
1	Введение в машинное обучение. Линейная алгебра. Библиотека NumPy. Метод ближайших соседей
2	Случайность. Наивный Байесовский классификатор
3	Оптимизация. Задача регрессии. Линейная регрессия
4	Обработка и визуализация данных. Матричные разложения
Модуль 3. Машинное обучение	
5	Задача классификации. Логистическая регрессия
6	Метод опорных векторов. Задача снижения размерности. Метод главных компонент
7	Решающие деревья и техники ансамблирования
8	Случайный лес. Продвинутое техники ансамблирования. Дилемма смещения-дисперсии
9	Градиентный бустинг
10	Оценка значимости признаков
11	Введение в глубокое обучение
Модуль 4. Глубокое обучение	
12	Основы глубокого обучения
13	Векторные представления слов
14	Рекуррентные нейронные сети. Проблема затухающего градиента
15	Обработка изображений. Сверточные нейронные сети



Форматы обучения

Асинхронный материал – предзаписанный контент, состоящий из видеолекций и конспектов для самостоятельного изучения.

Синхронные занятия – онлайн-семинары с преподавателем и группой. В неделю проходит 1 синхронная встреча длительностью 1,5 часа для каждой группы.

Форматы практических и самостоятельных заданий на курсе

В LMS:

- тест на оценку в конце каждой темы.

В Яндекс.Контесте*:

- 1-4 задания на программирование на оценку в конце каждой недели.

**оценки из Яндекс.Контеста переносятся в ЛМС еженедельно*

На онлайн-семинарах:

- короткая групповая работа для отработки навыков

Каникулы

На курсе есть каникулы: с 27 февраля по 3 марта и с 8 по 12 мая. В период каникул:

- не открываются новые материалы в ЛМС;
- не проводятся семинары.

Каникулы предусмотрены для того, чтобы компенсировать государственные праздники, выпадающие на семестр. Рекомендуем потратить время каникул на то, чтобы нагнать отставание, которое могло появиться, а также немного отдохнуть.

Оценивание

Формула оценки

Тесты в LMS*0,4 + задания на программирование*0,6

За что выставляются оценки:

- за прохождение **тестов** (по 1 каждые 3-4 недели);



- за решение **заданий на программирование** (по 1 на тему, т.е. каждые 1-2 недели).

Как проходит оценивание:

- **тесты** — после каждого прохождения LMS автоматически подсчитывает и сохраняет результат, на каждый тест дается **2 попытки**;
- **задания на программирование** — после каждого прохождения Яндекс.Контест автоматически подсчитывает и сохраняет результат, еженедельно баллы переносятся в ЛМС, на каждое задание дается **3 попытки**.

Дополнительные сложные задачи

По желанию вы можете решить две сложные задачи: **ML pipeline** и **Генерация поэзии**.

В ходе реализации ML pipeline вы получите опыт написания упорядоченного кода построения модели, который поможет вам в дальнейшем. А результатом работы над задачей “Генерация поэзии” станет ваш собственный **нейропоэт**.

Студентам, выполняющим дополнительные задачи, предоставляется возможность встретиться с семинаристом один на один для 15-минутной консультации.

Оценки за ML pipeline и Генерацию поэзии могут быть засчитаны вместо оценок за задания по некоторым неделям:

Сложная задача	За что засчитывается
ML pipeline	Задания на программирование по теме 3.2 “Метод опорных векторов. Оценка качества классификации. Методы кросс-валидации”
	Задания на программирование по теме 3.3 “Решающие деревья и техники ансамблирования”
	Задания на программирование по теме 3.4 “Случайный лес. Продвинутое техники ансамблирования. Дилемма смещения-дисперсии”
	Тестирование на неделе 9 “Метод опорных векторов. Оценка качества классификации. Методы кросс-валидации”
	Тестирование на неделе 12 “Случайный лес.



	Продвинутые техники ансамблирования. Дилемма смещения-дисперсии
Генерация Поэзии	Задания на программирование по теме 4.3 "Рекуррентные нейронные сети. Проблема затухающего градиента"
	Задания на программирование по теме 4.4 "Обработка изображений. Сверточные нейронные сети"
	Тестирование на неделе 22 Обработка изображений. Сверточные нейронные сети. Часть 2

К решению сложных задач допускаются не более 20 студентов из группы. Для допуска к решению задач необходимо пройти два этапа отбора:

- выполнить констест по Python на 70% до **08.03.2023 23:59 MSK**
- набрать в ЛМС 28% до **08.03.2023 23:59 MSK**

Студенты, сдавшие решение ML pipeline на любой балл, автоматически допускаются до решения Генерации поэзии, то есть пройти отбор необходимо только один раз за курс.

Если вы прошли отбор, мы оповестим вас по электронной почте или в telegram.

Дедлайны

На курсе предусмотрены строгие дедлайны: на каждое задание в LMS дается **2 недели** на выполнение (кроме недель 1-3, на них дается 3 недели). Дедлайны не переносятся.

За несоблюдение дедлайна снижается максимальный балл, который возможно получить за задание:

Соблюдение дедлайна	Максимальный балл
вовремя	100%
с опозданием до 2х недель	50%
с опозданием больше 2х недель	10%



Расписание дедлайнов указано в таблице ниже. Также дедлайн к каждому заданию установлен в ЛМС.

Неделя	Задания	Дедлайн
Модуль 2. Математические основы машинного обучения		
Неделя 1 Линейная алгебра. Библиотека NumPy 16.01 - 22.01	Задания на программирование по теме 2.1	13/02/2023 23:59 MSK
Неделя 2 Введение в машинное обучение. Метод ближайших соседей 23.01 - 29.01		
Неделя 3 Случайность. Наивный Байесовский классификатор 30.01 - 05.02	Задания на программирование по теме 2.2	20/02/2023 23:59 MSK
	Тестирование	20/02/2023 23:59 MSK
Неделя 4 Оптимизация 06.02 - 12.02	Задания на программирование по теме 2.3	27/02/2023 23:59 MSK
Неделя 5 Задача регрессии. Линейная регрессия 13.02 - 19.02		
Неделя 6 Обработка и визуализация данных. Матричные разложения 20.02 - 26.02	Задания на программирование по теме 2.4	06/03/2023 23:59 MSK
	Тестирование	06/03/2023 23:59 MSK
Неделя 7 КАНИКУЛЫ 27.02 - 05.03		
Модуль 3. Машинное обучение		
Неделя 8 Задача классификации. Логистическая регрессия 06.03 - 12.03	Задания на программирование по теме 3.1	20/03/2023 23:59 MSK



Неделя 9 Метод опорных векторов. Оценка качества классификации. Методы кросс-валидации 13.03 - 19.03	Задания на программирование по теме 3.2	27/03/2023 23:59 MSK
	Тестирование	27/03/2023 23:59 MSK
	ML pipeline	10/04/2023 23:59 MSK
Неделя 10 Решающие деревья и техники ансамблирования 20.03 - 26.03	Задания на программирование по теме 3.3	10/04/2023 23:59 MSK
Неделя 11 Seaborn и практика по деревьям 27.03 - 02.04		
Неделя 12 Случайный лес. Продвинутое техники ансамблирования. Дилемма смещения-дисперсии. 03.04 - 09.04	Задания на программирование по теме 3.4	17/04/2023 23:59 MSK
	Тестирование	17/04/2023 23:59 MSK
Неделя 13 Градиентный бустинг 10.04 - 16.04	Задания на программирование по теме 3.5	01/05/2023 23:59 MSK
Неделя 14 Градиентный бустинг на практике 17.04 - 23.04		
Неделя 15 Оценка значимости признаков 24.04 - 30.04	Задания на программирование по теме 3.6	08/05/2023 23:59 MSK
	Тестирование	08/05/2023 23:59 MSK
Неделя 16 Введение в глубокое обучение 01.05 - 07.05		
Неделя 17 КАНИКУЛЫ 08.05 - 14.05		
Модуль 4. Глубокое обучение		



Неделя 18 Оптимизация и регуляризация в глубоком обучении 15.05 - 21.05	Задания на программирование по теме 4.1	29/05/2023 23:59 MSK
Неделя 19 Векторные представления слов 22.05 - 28.05	Задания на программирование по теме 4.2	05/06/2023 23:59 MSK
Неделя 20 Рекуррентные нейронные сети. Проблема затухающего градиента 29.05 - 04.06	Задания на программирование по теме 4.3	12/06/2023 23:59 MSK
	Генерация поэзии	19/06/2023 23:59 MSK
Неделя 21 Обработка изображений. Сверточные нейронные сети. Часть 1 05.06 - 11.06	Задания на программирование по теме 4.4	26/06/2023 23:59 MSK
Неделя 22 Обработка изображений. Сверточные нейронные сети. Часть 2 12.06 - 18.06		
	Тестирование	26/06/2023 23:59 MSK

Критерии успешного прохождения курса

Этот курс является одновременно и дисциплиной вашей программы магистратуры, и программой дополнительного образования. Это значит, что в конце семестра у вас будет два контрольных этапа: в рамках высшего образования и в рамках дополнительного образования.

В рамках высшего образования за этот курс ставится **недифференцированный зачет**.

В рамках дополнительного образования вы имеете возможность получить **диплом о профессиональной переподготовке государственного образца** — официальный документ, который может дать преимущество при трудоустройстве.

Чтобы получить **зачет** во втором семестре, необходимо:

- набрать **не менее 30 баллов** суммарно за **LMS** до **22 мая 23:59 MSK**.

В зачетную книжку оценка ставится в недифференцированной форме (зачет/незачет).



Чтобы получить **диплом о профессиональной переподготовке государственного образца**, необходимо:

- иметь **зачет за первый семестр**;
- получить **зачет за второй семестр**;
- пройти **итоговый ассесмент** на платформе <https://unionepro.ru> с **18 по 26 июня**.

В диплом о профессиональной переподготовке оценка ставится в дифференцированной форме (отлично/хорошо/удовлетворительно). Вычисляется среднее арифметическое баллов за первый и второй семестр.

Соответствие баллов и оценки:

Балл, %	Оценка
80–100	Отлично
50-79	Хорошо
30-49	Удовлетворительно
0–29	Неудовлетворительно

Хакатон

В период с **19 по 30 июня 2023 г.** по программе будет проводиться хакатон. На хакатоне вы будете решать реальные задачи от заказчиков из IT-индустрии в командах.

За представление своего решения на хакатоне можно получить **до 20 баллов** к итоговому, и таким образом повысить свою оценку в дипломе о переподготовке.

Примеры расчета оценки

1. Дима получил 86 баллов (зачет) в первом семестре. Во втором семестре он набрал 75 баллов за тесты в LMS, 20 баллов за задания на программирование и прошел итоговый ассесмент. Дополнительные сложные задачи Дима не выполнял. На хакатон Дима пришел, но представить решение у его команды не получилось.

Результаты Димы:

Первый семестр: 86 баллов

Итоговый ассесмент:

Хакатон: (+0 баллов)

Балл за второй семестр: $75 \cdot 0,4 + 20 \cdot 0,6 = 42$

Итоговый балл: $(86 + 42) / 2 + 0 = 64$



Дима выполнил все условия получения зачета во втором семестре и диплома о переподготовке.

Оценка в зачетке: зачет

Оценка в дипломе о переподготовке: хорошо

2. Вика получила 49 баллов (зачет) в первом семестре. Во втором семестре она набрала 65 баллов за тесты в LMS, 85 баллов за задания на программирование. Команда Вики представила решение задачи на хакатоне, за него ребята получили 15 баллов. Итоговый ассесмент Вика пройти не успела.

Результаты Вики:

Первый семестр: 49 баллов

Итоговый ассесмент: ✗

Хакатон: ✓ (+15 баллов)

Балл за второй семестр: $65 \cdot 0,4 + 85 \cdot 0,6 = 77$

Итоговый балл: $(49 + 77) / 2 + 15 = 78$

Вика набрала проходной балл и поучаствовала в хакатоне, но не прошла обязательный ассесмент.

Оценка в зачетке: незачет

Без диплома о переподготовке

3. Алина получила 95 баллов (зачет) в первом семестре. Во втором семестре она набрала 30 баллов за тесты в LMS, 55 баллов за задания на программирование с учетом решения ML pipeline и Генерации поэзии. Команда Алины представила решение задачи на хакатоне и получила за него 20 баллов. Алина прошла итоговый ассесмент в установленные сроки.

Результаты Алины:

Первый семестр: 95 баллов

Итоговый ассесмент: ✓

Хакатон: ✓ (+20 баллов)

Балл за второй семестр: $30 \cdot 0,4 + 55 \cdot 0,6 = 45$

Итоговый балл: $(95 + 45) / 2 + 20 = 90$

Алина выполнила все условия получения зачета во втором семестре и диплома о переподготовке.

Оценка в зачетке: зачет

Оценка в дипломе о переподготовке: отлично