

# Машинное обучение и анализ данных

Лекторы:



**Язык:**

Русский

**Трудоемкость:**

3 з.е.

**Форма контроля:**

Зачет

**Образовательная программа:**

Теоретическая и экспериментальная физика

5 семестр

Беспроводные технологии

5 семестр

Беспроводные технологии (магистратура)

2 семестр

**Прerequisites:**

Первичная обработка данных

Выборочные характеристики

Проверка статистических гипотез

Прикладная статистика

Лекции (ак.час)*	Практические занятия (ак.час)	Лабораторные занятия (ак.час)
	16	
*1 академический час = 45 минутам		

Дисциплина «Машинное обучение (базовый уровень)» знакомит слушателей с задачами машинного обучения, решение которых востребовано в настоящее время. Подробно рассматривается задача регрессии и подходы к ее решению, задача классификации и некоторые метрические и статистические алгоритмы классификации: логистическая регрессия, метод k-ближайших соседей, байесовский классификатор, а также задача кластеризации с алгоритмами: k-средних и иерархическая (или агломеративная) кластеризация. Дисциплина также рассказывает о более сложных и продвинутых алгоритмах МО для решения различных задач профессиональной области. Он знакомит и с элементами факторного анализа, и с деревьями принятия решений, и ансамблевыми методами. Кроме того, рассматриваются еще и такие ветки машинного обучения, как обучение с подкреплением и нейронные сети.

## Содержание курса

### 5 семестр

#### Машинное обучение и анализ данных

##### Структура курса

Разделы	Консультации (ак.ч.)
1.1. Введение в машинное обучение	16
1.2. Элементы языка программирования Python: переменные, типы и структуры данных	
1.3. Элементы языка программирования Python: библиотеки matplotlib, numpy, pandas	
1.4. Инструменты машинного обучения	
1.5. Метод главных компонент и его применения к визуализации, уменьшению размерности, сжатию	
1.6. Задача регрессии, метод наименьших квадратов, множественная регрессия, полиномиальная регрессия	
1.7. Статистические характеристики параметров регрессии	
1.8. Задача классификации. Метрический подход и метод k-пп, выбор метрики, выбор числа k, обобщения k-пп	
1.9. Задача классификации. Вероятностный подход: наивный байесовский классификатор и его обобщения	
1.10. Задача классификации. Логистическая регрессия, многоклассовое обобщение	
1.11. F-мера и ROC анализ	
1.12. Метод опорных векторов в случае линейно разделкой и линейно неразделимой выборки. Ядра и kernel trick	
1.13. Понятие энтропии и условной энтропии. Прирост информации	
1.14. Задача классификации. Деревья принятия решений	
1.15. Задача кластеризации. Типы кластеров, метод K-средних, каменная осыпь, начальная инициализация кластеров	
1.16. Задача кластеризации. Агломеративная кластеризация и DBSCAN, особенности реализации	
1.17. Ресемплинг: джекнайф и бутстрап, статистические характеристики	
1.18. Ансамбли моделей и их использование в машинном обучении	
1.19. Обучение с подкреплением: многорукие бандиты	
1.20. Обучение с подкреплением: алгоритмы SARSA и Q-learning	

##### Рекомендуемые ресурсы

1. Рашка, С. Python и машинное обучение [Электронный ресурс] : рук. / С. Рашка ; пер. с англ. Логунова А.В.. — Электрон. дан. — Москва : ДМК Пресс, 2017. — 418 с. — Режим доступа: <https://e.lanbook.com/book/100905>
2. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] — Электрон. дан. — Москва : ДМК Пресс, 2015. — 400 с. — Режим доступа: <https://e.lanbook.com/book/69955>
3. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262> (дата обращения: 20.01.2023).
4. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2018. — 490 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Режим доступа: <https://urait.ru/book/analiz-dannyh-412967>
5. Машинное обучение / Университет ИТМО — Открытое образование. — Режим доступа: <https://openedu.ru/>

##### Политика оценивания

## **Оценочные средства дисциплины: итоговое упражнение, рубежный зачет, итоговый зачет.**

**1. Итоговое упражнение.** После каждой лекции студентам выдаются индивидуальные задачи. Задачи включают в себя некоторое количество (1-20) заданий. Задание выполняется студентом при помощи сети Интернет и инструментов, рассматриваемых в лекции. Срок выполнения – не позднее срока освоения дисциплины. Для ввода ответа на задания дается от двух до десяти попыток. Подпункты задачи оцениваются равномерно. Если в задаче N полей ввода, то каждое оценивается согласно формуле  $60/M/N$ , где M — количество задач в курсе (от 10 до 16). **Итого, максимально возможный балл за задачу равен 60/M. Минимальный балл по задаче не предусмотрен (равен 0)**

**2. Рубежный зачет.** Студент получает индивидуальные задачи по лекциям первой половины курса. Задачи включают в себя некоторое количество (1-20) заданий. Тестовое задание выполняется студентом при помощи сети Интернет и инструментов, рассматриваемых в лекции за ограниченное время (60-90 минут). Для проверки ответа на задания дается от двух до пяти попыток. Попытка сдачи зачета одна. Зачет является ключевой точкой.

Варианты задач генерируются автоматически и представляют собой задачи как с закрытым, так и открытым типом вопросов. Каждое задание оценивается в  $20/N$  баллов, где N — количество заданий, **20 — максимальное количество баллов за контрольную работу. Минимальное количество баллов для получения зачета по контрольной работе равно 10.** Тестирование считается выполненным, если студент ответил верно на  $N/2$  задний (т.е. набрал 50% или более).

**3. Итоговый зачет.** Студент получает индивидуальные задачи по лекциям первой половины курса. Задачи включают в себя некоторое количество (1-20) заданий. Тестовое задание выполняется студентом при помощи сети Интернет и инструментов, рассматриваемых в лекции за ограниченное время (60-90 минут). Для проверки ответа на задания дается от двух до пяти попыток. Попытка сдачи зачета одна. Зачет является ключевой точкой.

Варианты задач генерируются автоматически и представляют собой задачи как с закрытым, так и открытым типом вопросов. Каждое задание оценивается в  $20/N$  баллов, где N — количество заданий, **20 — максимальное количество баллов за контрольную работу. Минимальное количество баллов для получения зачета по контрольной работе равно 10.** Тестирование считается выполненным, если студент ответил верно на  $N/2$  задний (т.е. набрал 50% или более).

**Для зачета по дисциплине необходимо выполнить индивидуальные задачи курса, обязательный рубежный зачет (порог 50%), итоговый зачет (порог 50%) и набрать в сумме не менее 60 баллов.**

Знания, умения и навыки обучающихся при промежуточной аттестации в форме зачета определяются оценками «зачтено», «не зачтено».

«Зачтено» – обучающийся знает курс на уровне материала онлайн-курса, умеет привести разные точки зрения по излагаемому вопросу.

«Не зачтено» – обучающийся имеет пробелы в знаниях основного учебного материала, допускает принципиальные ошибки в выполнении предусмотренных программой заданий.