

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Суворов Антон Дмитриевич  
Должность: Ректор  
Дата подписания: 13.06.2025 20:30:47  
Уникальный программный ключ:  
a39bdb15d680d5b0adb1cedda75c1efb14747dcd

# СКОЛКОВСКИЙ ИНСТИТУТ НАУКИ И ТЕХНОЛОГИЙ (Сколтех)

|                              |                   |
|------------------------------|-------------------|
| Рабочая программа дисциплины | Машинное обучение |
|------------------------------|-------------------|

|               |   |
|---------------|---|
| Преподаватель | Бурнаев Евгений Владимирович, профессор, д.ф-м.н., доцент |
|---------------|---|

## Аннотация

### Описание курса

Курс представляет собой общее введение в машинное обучение (МЛ) и его приложения. Курс охватывает фундаментальные темы в МЛ и описывает наиболее важные алгоритмические основы и подходы, аспекты применения алгоритмов. Курс начинается с обзора канонических приложений и задач МЛ, сценариев обучения и т.д. Далее, в курсе подробно обсуждаются фундаментальные алгоритмы МЛ для классификации, регрессии, кластеризации и т.д., их свойства и практическое применение. Последняя часть курса посвящена продвинутым темам МЛ, таким как гауссовские процессы, нейронные сети. В практических сессиях мы покажем, как использовать методы МЛ и настраивать их гиперпараметры. Домашние задания включают в себя применение существующих алгоритмов для решения задач анализа данных. Предполагается, что студенты, посещающие данный курс, уже знакомы с основными понятиями линейной алгебры, теории вероятностей, математического анализа, оптимизации и программирования на python.

По окончании курса студенты будут:

- Иметь хорошее понимание фундаментальных проблем и проблем МЛ: анализ свойств данных, выбор типа модели и сложности соответствующего функционального класса, и т.п.;
- Иметь представление о сильных и слабых сторонах многих популярных подходов МЛ;
- Понимать лежащие в основе МЛ математические концепции, позволяющие проводить обучение с учителем и обучение без учителя;
- Уметь разрабатывать и использовать алгоритмы машинного обучения для решения прикладных задач.

## Основная информация

|                             |                             |
|-----------------------------|-----------------------------|
| Академический уровень курса | Магистратура<br>Аспирантура |
| Количество кредитов         | 6                           |

## Предварительные требования к курсу / рекомендации

Обязательными предметами курса являются численная линейная алгебра, математический анализ, теория вероятностей, оптимизация и программирование (python).

Тип оценки - дифференцированная

Отображение оценок в процентах

|           |    |
|-----------|----|
| <b>A:</b> | 86 |
| <b>B:</b> | 76 |
| <b>C:</b> | 66 |
| <b>D:</b> | 56 |
| <b>E:</b> | 46 |
| <b>F:</b> | 0  |

## 1. Содержание курса

| Тема                    | Краткое содержание  | Лекции (час) | Семинары (час) | Лабораторные занятия (час) | Самостоятельная работа (час) |
|-------------------------|---|--------------|----------------|----------------------------|------------------------------|
| Вводная лекция          | - Введение. Некоторые канонические приложения и проблемы. Определения и терминология<br>- Типы проблем: Контролируемые, полуконтролируемые, эмпирические риски<br>Минимизация, перекрестная проверка  | 1,5          | 1,5            |                            | 6                            |
| Регрессия, трюк с ядром | - Регрессия (постановка задачи)<br>- Линейная регрессия. Закрытая форма Решение Регрессия гребня. Замкнутая форма Решение. Прямое двойственное решение<br>- Нелинейный случай. Ядра. Хребет ядра Регрессия<br>- LASSO. Разреженность L1.<br>Эластичная сеть   | 1,5          | 1,5            |                            | 6                            |
| Классификация           | - Бинарная классификация, кривая потерь, ROC / AUC, точность и отзыв<br>- Изучение классификатора. Заменитель Минимизация эмпирического риска потерь (ERM), Переоснащение.<br>- Регуляризация, логарифмические потери<br>- Двухклассовая и многоклассовая логистика Регрессия<br>- k-Классификатор ближайших соседей. | 1,5          | 1,5            |                            | 6                            |

|  |  |     |     |  |   |
|--|--|-----|-----|--|---|
|  | <p>Гипотезы компактности и непрерывности. Функции расстояния</p> <ul style="list-style-type: none"> <li>- Деревья классификации и регрессии.</li> </ul> <p>Ансамбли</p> <ul style="list-style-type: none"> <li>- Наивный байесовский классификатор</li> </ul>  |     |     |  |   |
| <p>Древовидные методы.<br/>Случайный лес</p>   | <ul style="list-style-type: none"> <li>- Классификаторы дерева решений. Алгоритм разделения и покорения и алгоритм ID3.</li> <li>- Жадное дерево для классификации.</li> <li>- Варианты проектирования для изучения дерева решений: выбор корня; чистота, энтропия, прирост информации и индекс Джини; функция потерь.</li> <li>- Обрезка и переоснащение: до и после обрезки.</li> </ul>  | 1,5 | 1,5 |  | 6 |
| <p>Адабуст</p>   | <ul style="list-style-type: none"> <li>- Ансамбли классификаторов. Объединение в пакеты.</li> <li>- Многоуровневое обобщение.</li> <li>- Повышение эффективности. Мотивация. Пример фильтрации спама: подход, основанный на повышении голосов большинством голосов. Повышение для бинарной классификации. Повышение эвристики: непрерывные и экспоненциальные верхние границы.</li> <li>- AdaBoost. Примеры с игрушками и проблемы с шумом. Улучшаем пни.</li> </ul>   | 1,5 | 1,5 |  | 6 |
| <p>Повышение градиента</p>   | <ul style="list-style-type: none"> <li>- Повышение градиента. Механизм повышения градиента (GBM). GBM для регрессии. Регуляризация GBM с помощью сжатия. Теоретическая мотивация.</li> <li>- Деревья решений для повышения градиента. Дерево Ансамбли.</li> </ul>  | 1,5 | 1,5 |  | 6 |
| <p>Расширенная классификация<br/>Случаи несбалансированности и множественно й маркировки</p> | <ul style="list-style-type: none"> <li>- Несбалансированная классификация. Коэффициент несбалансированности.</li> <li>- Методы повторной выборки: случайная избыточная выборка (ROS); случайная недостаточная выборка (RUS); метод синтетической неполной выборки (SMOTE).</li> <li>- Классификация по нескольким классам. Один против всех. Один на один. Подход, основанный на ошибке</li> <li>- Корректирующие коды.</li> <li>- Многоклассовые алгоритмы: Логистические Регрессия, SVMs. Байесовский подход.</li> <li>- Дополнительно: Непараметрическая оценка.</li> </ul> | 1,5 | 1,5 |  | 6 |

|                                     |  |     |     |  |   |
|-------------------------------------|--|-----|-----|--|---|
|                                     | <p>Средняя интегральная квадратическая ошибка (MISE).<br/>Избыточное сглаживание, недостаточное сглаживание.<br/>- Дополнительно: Оценка плотности ядра (KDE). Многомерный KDE.<br/>- Дополнительно: непараметрическая регрессия.<br/>Оценка Надарая-Уотсона.</p>  |     |     |  |   |
| Модель и особенность выбора         | <p>- Переобучение. Исправление ошибок. Предвзятость-<br/>Выбор оптимального варианта.<br/>- Гиперпараметры. Выбор модели.<br/>Ошибки в обучении, тестировании и валидации. Пересекать-<br/>Утверждение. Согласованность модели.<br/>- Выбор элемента. Функция "ЛАССО" Выбор.<br/>- Регуляризация. Байесовский взгляд на Регуляризацию.<br/>- Статистика Ср Маллоу. AIC. BIC.<br/>- Дополнительно: Анализ чувствительности. Элементарные эффекты. Индексы Соболя. EASI и CSTA.</p>  | 1,5 | 1,5 |  | 6 |
| Мелкий Искусственный Нервный Сети   | <p>- Мотивация. Нейроны из реальной жизни.<br/>- Алгоритм персептрона и SGD.<br/>Конвергенция.<br/>- Нейронные сети с прямой связью.<br/>Функция активации.<br/>- Нейронные сети для классификации. Потеря логарифма.<br/>- Прямое и обратное распространение<br/>Алгоритм обратной обработки.<br/>Проблемы.</p>   | 1,5 | 1,5 |  | 6 |
| Глубокий Искусственный Нервный Сети | <p>- Олдскульный подход к разработке объектов.<br/>Магия построения объектов.<br/>- Нейронная сеть как вычислительный граф.<br/>- Примеры практического применения. Причины глубокого прорыва в NN<br/>- Универсальная аппроксимация<br/>- Обучение вычислительных графов с помощью обратной обработки.<br/>- Представление изображений (RGB). Сверточные слои.<br/>Объединение слоев в пул.<br/>- AlexNet. VGG. Inception V3.<br/>ResNet.<br/>- Передача знаний посредством тонкой настройки.<br/>- Рекуррентные сети (кратко).<br/>- Пакетная норма. Норма веса.<br/>Выпадение. Увеличение объема данных</p> | 1,5 | 1,5 |  | 6 |
| Нейронные                           | - Рекуррентные нейронные сети  | 1,5 | 1,5 |  | 6 |

|  |  |     |     |  |   |
|--|--|-----|-----|--|---|
| сети для обработки последовательных данных | <ul style="list-style-type: none"> <li>- LSTM и GRU</li> <li>- Основная идея внимания</li> <li>- Внимание к себе</li> <li>- Преобразователи для последовательных данных</li> <li>- Визуальные преобразователи</li> <li>- Чувствительность и внимание</li> </ul>  |     |     |  |   |
| Гауссовский Процесс                        | <ul style="list-style-type: none"> <li>- Байесовское моделирование. Стохастический Процесс. Случайная функция.</li> <li>- Определение гауссовского процесса. Совместный Гауссово распределение. Среднее значение и ковариационная функция.</li> <li>- Гауссово ядро как ковариационная функция. Интерпретируемость ядра</li> <li>Параметры.</li> <li>- Арифметика ядра. Построение Новых ядер.</li> <li>- Графическая регрессия с шумом: Модель. Предсказание.</li> <li>Интерполяция. Сглаживающий.</li> <li>- Оптимизация GP.</li> <li>- Классификация GP.</li> <li>Сигмоидальное правдоподобие.</li> <li>- Нестационарный GP.</li> </ul> | 1,5 | 1,5 |  | 6 |
| Уменьшение размеров                        | <ul style="list-style-type: none"> <li>- Постановка задачи. Примеры: Лица, профили, MNIST.</li> <li>- PCA</li> <li>- Многомерное масштабирование (MDS)</li> <li>- Репликативные нейронные сети (Автоэнкодеры)</li> <li>- Граф, основанный на ближайших соседях: ISOMAP, LLE</li> <li>- t-стохастическое вложение соседей</li> </ul>  | 1,5 | 1,5 |  | 6 |
| Аномалия Обнаружение                       | <ul style="list-style-type: none"> <li>- Обнаружение аномалий.</li> <li>- Методы, основанные на ближайших соседях</li> <li>- SVM одного класса. Выбор ядра</li> <li>- Другие подходы к обнаружению аномалий</li> </ul>   | 1,5 | 1,5 |  | 6 |
| Кластеризация                              | <ul style="list-style-type: none"> <li>- Иерархическая кластеризация (агломеративные/разделяющие модели).</li> <li>- K-средние значения</li> <li>- Валидность кластера. Внешние показатели: Энтропия, Взаимная информация, Jaccard Индекс, Rand Index, Silhouette Коэффициент</li> <li>Модели смешивания и т.д.</li> <li>- Жесткое и мягкое присвоение значений с помощью Kmeans, гауссовых моделей смешения (GMM)</li> <li>- Алгоритм максимизации математического ожидания.</li> <li>Конвергенция EM. K-средние значения в сравнении с обучающей GMM</li> </ul>  | 1,5 | 1,5 |  | 6 |

|                         |  |  |    |   |  |
|-------------------------|--|--|----|---|--|
| Консультации по проекту | Консультации по окончательным групповым проектам.  |  | 18 |   |  |
| Контактные часы         | Дополнительные часы для связи в случае возникновения дополнительных вопросов от студентов. |  |    | 9 |  |

## 2. Результаты обучения

Результаты обучения в Сколтехе указаны в соответствии со структурой результатов обучения в Сколтехе

### 1. ФУНДАМЕНТАЛЬНЫЕ ЗНАНИЯ

1.1. Знание математики и естественных наук

1.2. Знание прикладной науки и техники, науки, в том числе современные методы и инструменты

### 2.1. ПОЗНАНИЕ И СПОСОБЫ РАССУЖДЕНИЯ

2.1.1. Аналитическое мышление и решение проблем

2.1.5. Критическое мышление и метапознание

### 2.2. ОТНОШЕНИЕ И ПРОЦЕСС ОБУЧЕНИЯ

2.2.3. Ответственность, интенсивность, настойчивость, безотлагательность и воля к достижению поставленных целей

2.2.4. Находчивость, гибкость и способность к адаптации

2.2.6. Развитие и поддержка преподавательского состава и обучающегося сообщества

### 2.3. ЭТИКА, СПРАВЕДЛИВОСТЬ И ДРУГИЕ ОБЯЗАННОСТИ

2.3.1. Этические действия, честность и мужество

### 3.1. КОММУНИКАЦИЯ В МЕЖДУНАРОДНОЙ СРЕДЕ

3.1.2. Письменная, электронная и графическая коммуникация

3.1.3. Устная презентация и обсуждение

3.1.4. Вопросы, слушание и диалог

3.1.5. Общение на английском языке в научной, деловой и общественной среде

### 3.2. КОМАНДНАЯ РАБОТА И ЛИДЕРСТВО

3.2.1. Формирование эффективных команд

3.2.2. Командная работа и управление проектами

### 3.3. СОТРУДНИЧЕСТВО И ИЗМЕНЕНИЯ

3.3.1. Установление разнообразных связей и сетевого взаимодействия

### 3.3. СОТРУДНИЧЕСТВО И ИЗМЕНЕНИЯ

3.3.3. Переговоры и разрешение конфликтов

## 3. Задания и выставление оценок

|   |    |
|---|----|
| Требование к физической посещаемости (% от числа занятий) | 80 |
|---|----|

К требованиям о личном посещении:

студентам разрешается пропускать занятия только по уважительной причине, имеющей обоснование, например, медицинское заключение

| Тип назначения        | Краткое содержание задания   | % от итоговой оценки за курс |
|-----------------------|--|------------------------------|
| Домашние задания      | В этом домашнем задании вам нужно выполнить некоторые задачи с помощью классических пакетов машинного обучения на Python, таких как NumPy, Matplotlib и т.д., а также реализовать и сравнить основные алгоритмы машинного обучения с контролем   | 15                           |
| Домашние задания      | В этом домашнем задании вам нужно решить несколько задач, связанных с нейронными сетями, гауссовыми процессами и алгоритмами машинного обучения без контроля.  | 15                           |
| Домашние задания      | В этом задании вам необходимо создать тестовые вопросы на основе содержания лекции.  | 5                            |
| Финальный проект      | Предварительный вариант отчета по проекту. В начале семестра преподаватель представляет студентам набор проектов. Студенты объединяются в группы. Затем они будут работать над выбранным проектом. Проекты могут включать в себя повторные исследования, новые работы или аналитические задания. Студенты должны предложить более ранний вариант итогового проекта с указанием цели проекта, основных задач и обзора литературы. | 5                            |
| Финальный экзамен     | Экзамен состоит из нескольких теоретических вопросов, в основном с множественным выбором.  | 20                           |
| Финальный проект      | Краткий отчет о проекте с 4-6 пунктами страницы и репозиторий на github с полностью воспроизводимым кодом. По истечении крайнего срока подачи проекта учащиеся должны будут представить его в классе. Окончательная версия проекта, кода и презентации будет оценена.  | 30                           |
| Промежуточный экзамен | Промежуточный экзамен состоит из нескольких теоретических вопросов, в основном в формате множественного выбора, охватывающих первую половину содержания курса.   | 10                           |

#### 4. Критерии оценки

|                       |                  |
|-----------------------|------------------|
| <u>Задание 1 Типа</u> | Домашние задания |
|-----------------------|------------------|

### Пример задания 1

[https://docs.google.com/document/d/19ri1FwpCK4nKp32IUrAVKj12\\_zn6YOvVn\\_z7xzjmiKcg/edit?usp=sharing](https://docs.google.com/document/d/19ri1FwpCK4nKp32IUrAVKj12_zn6YOvVn_z7xzjmiKcg/edit?usp=sharing)

Критерии оценки:

Стоимость домашних заданий составляет 35% от итоговой оценки. Каждое домашнее задание состоит из нескольких задач по кодированию. Каждая задача (или ее итоговый балл) оценивается как 1, если задача полностью решена (предоставлен полностью воспроизводимый код, результаты оценки сохранены и корректны, необходимые комментарии предоставлены студентом), 0 - в противном случае. После каждого домашнего задания будет проводиться офлайн-тест, основанный на содержании домашнего задания, для проверки оригинальности и предотвращения плагиата.)

|                              |                  |
|------------------------------|------------------|
| <b><u>Задание 2 Типа</u></b> | Финальный проект |
|------------------------------|------------------|

### Пример задания 2

Краткий отчет о проекте на 4-6 страницах, репозиторий на github с полностью воспроизводимым кодом. По истечении крайнего срока подачи проекта студенты должны будут представить проект в классе.

Основные критерии оценки:

5/35 - (групповая) оценка за качество предварительного отчета

20/35 - (групповая) оценка за качество окончательного проекта

10/35 - (групповая) оценка за качество презентации

|                              |                   |
|------------------------------|-------------------|
| <b><u>Задание 3 Типа</u></b> | Финальный экзамен |
|------------------------------|-------------------|

### Пример задания 3

Они состоят из нескольких теоретических вопросов, в основном с множественным выбором.

### Критерии оценки для задания 3

Критерии оценки для итогового экзамена направлены на оценку точности, ясности, структуры и глубины ответов учащихся. Чтобы получить максимальную оценку, учащиеся должны предоставить правильные и хорошо структурированные ответы, демонстрирующие четкое понимание материала курса, логическое мышление и способность эффективно применять концепции. Итоговый экзамен охватывает все содержание курса.

**Задание 4 Типа**

Промежуточный экзамен

**Пример задания 4**

Промежуточный экзамен состоит из нескольких теоретических вопросов, в основном в формате множественного выбора, охватывающих первую половину содержания курса.

**Критерии оценки для задания 4**

Критерии оценки для промежуточной аттестации, аналогичные критериям для экзамена, подчеркивают точность, ясность, структуру и глубину ответов учащихся. Чтобы получить наивысшую оценку, учащиеся должны предоставить точные и хорошо организованные ответы, демонстрирующие глубокое понимание материала, логическое мышление и способность эффективно применять концепции. В ходе промежуточной аттестации основное внимание уделяется содержанию первой половины курса.

**5. Учебники и интернет-ресурсы**

| Необходимые учебники   | ISBN-13 (or ISBN-10) |
|--|----------------------|
| Bishop, C.M. Pattern Recognition and Machine Learning. Springer, 2007  | 9780387310732        |
| Barber, D. Bayesian Reasoning and Machine Learning. Cambridge University Press, 2012   | 9780521518147        |
| Рекомендуемые учебники   |                      |
| Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009. | 9780387848587        |
| Shai Shalev-Shwartz, Shai Ben-David. Understanding Machine Learning: From Theory to Algorithms. Cambridge, 2014.                                 | 9781107057135        |
| R.E. Schapire, Y. Freund. Boosting. MIT, 2012  | 9780262526036        |
| M. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of Machine Learning. MIT, 2012.   | 9780262018258        |
| B. Clarke, E. Fokoue, H.H. Zhang. Principles and Theory for Data Mining and Machine Learning. Springer, 2009                                     | 9780387981352        |
| Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.   | 9780262018029        |
| Sutton and Barto. Reinforcement Learning: An Introduction. MIT Press, 1998.  | 9780262193986        |

| Веб-ресурсы (ссылки)  | Описание   |
|---|--|
| <a href="http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage">http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage</a> | онлайн-версия книги Барбера "Байесовские рассуждения и машина Обучение"                                |
| <a href="http://gaussianprocess.org/gpml/">http://gaussianprocess.org/gpml/</a>   | Carl Rasmussen and Christopher Williams. Gaussian Processes for Machine Learning. The MIT Press, 2006. |
| <a href="http://wol.ra.phy.cam.ac.uk/mackay/itila/book.html">http://wol.ra.phy.cam.ac.uk/mackay/itila/book.html</a>   | D. J. C. MacKay (2003) Information Theory, Inference, and Learning Algorithms.                         |

## 6. Оборудование

|                         |
|-------------------------|
| Программное обеспечение |
| Google Colab            |