

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы бизнеса
высоких технологий**

В.Ю. Григорьев

	Рабочая программа дисциплины (модуля)
по дисциплине:	Машинное обучение
по направлению:	Наукоёмкие технологии и экономика инноваций
профиль подготовки:	Создание и развитие высокотехнологичного бизнеса Физтех-школа бизнеса высоких технологий Физтех-школа бизнеса высоких технологий
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: П.И. Ахтямов, старший преподаватель

Программа обсуждена на заседании Физтех-школы бизнеса высоких технологий 07.04.2024

Аннотация

Курс “Машинное обучение” охватывает широкий спектр тем, начиная от основ машинного обучения и заканчивая глубокими нейросетями и методами обучения без учителя. В ходе обучения студенты познакомятся с различными алгоритмами и моделями машинного обучения, научатся применять их на практике и анализировать полученные результаты.

Курс предназначен для студентов, интересующихся машинным обучением и желающих освоить современные методы и алгоритмы машинного обучения.

1. Цели и задачи

Цель дисциплины

- познакомить студентов с основными концепциями и методами машинного обучения;
- научить студентов применять методы машинного обучения для решения практических задач;
- сформировать у студентов понимание основных проблем и ограничений машинного обучения.

Задачи дисциплины

- научиться оценивать качество моделей машинного обучения;
- научиться применять модели машинного обучения для решения практических задач;
- изучить основные библиотеки и инструменты для машинного обучения.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-5 Способен собирать, обрабатывать и интерпретировать с использованием современных информационных технологий данные, необходимые для профессиональной сферы деятельности	ОПК-5.1 Владеет знаниями и навыками использования информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-5.2 Способен применить знание информационно-коммуникационных технологий для решения поставленной задачи, формулирования выводов и оценки полученных результатов
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-10 Способен применять методы планирования исследований и экспериментов при выполнении проектов и заданий в избранной предметной области	ПК-10.3 Владеет методами планирования исследований и экспериментов в избранной предметной области
	ПК-10.2 Умеет применять теоретические знания к построению программ исследований и экспериментов при выполнении конкретных проектов и заданий

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные понятия и определения машинного обучения;
- типы задач машинного обучения и методы их решения;
- методы предварительной обработки данных для машинного обучения;
- принципы построения и настройки моделей машинного обучения.

уметь:

- применять методы машинного обучения для решения практических задач;
- строить и настраивать модели машинного обучения;
- оценивать качество моделей машинного обучения;
- работать с библиотеками и инструментами для машинного обучения.

владеть:

- навыками построения и настройки моделей машинного обучения;
- навыками оценки качества моделей машинного обучения;
- навыками работы с библиотеками и инструментами для машинного обучения;
- навыками анализа результатов машинного обучения и формулирования выводов.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение в машинное обучение	2	2		3
2	Линейные модели	4	4		6
3	Задачи классификации	4	4		6
4	Нелинейные методы классификации	2	2		3
5	Деревья решений	2	2		3
6	Ансамблевые подходы	4	4		6
7	Базовые нейросетевые подходы	2	2		3
8	Базовая обработка текстов	2	2		3
9	Рекуррентные нейронные сети	2	2		3
10	Сверточные нейронные сети	4	4		6
11	Обучение без учителя	2	2		3
Итого часов		30	30		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 1 (Осенний)

1. Введение в машинное обучение

Задача машинного обучения. Понятие объекта, признака, целевой переменной. Функция потерь, эмпирическая функция риска, метрика качества. Обучающая и валидационная выборки, тестовая выборка. Виды задач машинного обучения.

2. Линейные модели

Формулировка задачи. Метрики для решения задачи регрессии: MSE, MAE, MAPE, SMAPE. Регуляризация моделей.

3. Задачи классификации

Пороговая функция потерь. Понятие отступа. Сведение задачи классификации к градиентным методам. Градиентный метод и стохастический градиентный метод: преимущества и недостатки. Логистическая регрессия.

4. Нелинейные методы классификации

Методы сжатия размерности: метод главных компонент (PCA). Метод ближайших соседей, применение ядерных функций.

5. Деревья решений

Концепция дерева решений. Критерии информативности: критерий информативности, критерий Джини. Применение деревьев решений для задач классификации и регрессии. Гиперпараметры деревьев решений: высота дерева, количество объектов в листе.

6. Ансамблевые подходы

Подход Bootstrap Aggregation, Boosting. Случайный лес - преимущества и недостатки. Градиентный Boosting, AdaBoost.

7. Базовые нейросетевые подходы

Композиция линейных моделей. Функции активации. Прямое распространение. Метод обратного распространения ошибки (BackPropagation). Проблема затухания градиента.

8. Базовая обработка текстов

Представление текстовых видов - мешок слов, латентное представление. Предобработка текстов - лемматизация, стемминг. Подходы Word2Vec, fasttext. Количественные представления - TD-IDF, "мешок слов".

9. Рекуррентные нейронные сети

Представление форматов последовательностей. Рекуррентные нейронные сети. Метод обратного распространения ошибки по времени. Представление рекуррентной нейронной сети в виде марковских цепей. Слои LSTM, GRU.

10. Сверточные нейронные сети

Отличие сверточного слоя от линейного слоя. Слои MaxPooling и average pooling. Концепция "головы" и "хвоста" в сверточной нейронной сети. Архитектуры ResNet, Inception, VGG.

11. Обучение без учителя

Метрики качества - ARI, AMI, оценка гомогенности. Методы кластеризации - DBScan, иерархическая кластеризация. Методы визуализации - PCA, t-SNE, UMAP.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

6. Перечень рекомендуемой литературы

Основная литература

1. Введение в машинное обучение с помощью Python%Introduction to Machine Learning with Python, руководство для специалистов по работе с данными / А. Мюллер, С. Гвидо, Москва ; Санкт-Петербург, Диалектика, 2019
2. Теория вероятностей и математическая статистика / В. Е. Гмурман. – Москва: Юрайт, 2022.

Дополнительная литература

1. Машинное обучение с подкреплением на Python /Министерство науки и высшего образования Российской Федерации, Московский физико-технический институт (национальный исследовательский университет), Кафедра системных исследований ; составители: А. И. Панов, А. А. Скрынник , Москва, МФТИ, 2019

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Не используются

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

На занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике.

Успешное освоение дисциплины требует:

- посещения студентом всех видов аудиторных занятий;
- ведения конспекта в ходе лекционных занятий;
- качественной самостоятельной подготовки к практическим занятиям, активной работы на них;
- активной самостоятельной и аудиторной работы студента;
- своевременной сдачи преподавателю заданий по аудиторным видам работ.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Научноёмкие технологии и экономика инноваций
профиль подготовки:	Создание и развитие высокотехнологичного бизнеса Физтех-школа бизнеса высоких технологий Физтех-школа бизнеса высоких технологий
курс:	<u>1</u>
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

Разработчик: П.И. Ахтямов, старший преподаватель

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-5 Способен собирать, обрабатывать и интерпретировать с использованием современных информационных технологий данные, необходимые для профессиональной сферы деятельности	ОПК-5.1 Владеет знаниями и навыками использования информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
	ОПК-5.2 Способен применить знание информационно-коммуникационных технологий для решения поставленной задачи, формулирования выводов и оценки полученных результатов
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.1 Способен находить, анализировать и обобщать информацию об актуальных результатах исследований в рамках тематической области своей профессиональной деятельности
	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
ПК-10 Способен применять методы планирования исследований и экспериментов при выполнении проектов и заданий в избранной предметной области	ПК-10.3 Владеет методами планирования исследований и экспериментов в избранной предметной области
	ПК-10.2 Умеет применять теоретические знания к построению программ исследований и экспериментов при выполнении конкретных проектов и заданий

2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение» обучающийся должен:

знать:

- основные понятия и определения машинного обучения;
- типы задач машинного обучения и методы их решения;
- методы предварительной обработки данных для машинного обучения;
- принципы построения и настройки моделей машинного обучения.

уметь:

- применять методы машинного обучения для решения практических задач;
- строить и настраивать модели машинного обучения;
- оценивать качество моделей машинного обучения;
- работать с библиотеками и инструментами для машинного обучения.

владеть:

- навыками построения и настройки моделей машинного обучения;
- навыками оценки качества моделей машинного обучения;
- навыками работы с библиотеками и инструментами для машинного обучения;
- навыками анализа результатов машинного обучения и формулирования выводов.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

1. Введение в ML
2. Линейные модели.
3. Задачи классификации.
4. Нелинейные методы классификации.
5. Деревья решений.
6. Ансамблевые подходы.

7. Базовые нейросетевые подходы.
8. Базовая обработка текстов.
9. Рекуррентные нейронные сети.
10. Сверточные нейронные сети.
11. Обучение без учителя.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Что такое машинное обучение и какие задачи оно решает?
2. Какие существуют типы задач машинного обучения и как они отличаются друг от друга?
3. Что такое функция потерь и как она используется в машинном обучении?
4. Что такое эмпирическая функция риска и как она связана с функцией потерь?
5. Что такое метрика качества и как она используется для оценки эффективности моделей машинного обучения?
6. Что такое обучающая, валидационная и тестовая выборки и как они используются в машинном обучении?
7. Какие существуют метрики для решения задачи регрессии и как они работают?
8. Что такое регуляризация моделей и зачем она нужна?
9. Что такое пороговая функция потерь и как она используется в задачах классификации?
10. Что такое градиентный метод и стохастический градиентный метод и как они применяются в машинном обучении?
11. В чём заключается метод главных компонент (PCA) и как он используется для сжатия размерности данных?
12. Что такое метод ближайших соседей и как он применяется для классификации данных?
13. Что такое деревья решений и как они работают?
14. Какие существуют критерии информативности для деревьев решений и как они выбираются?
15. Что такое ансамблевые подходы и как они улучшают точность моделей машинного обучения?
16. В чём заключается подход Bootstrap Aggregation и как он работает?
17. Что такое случайный лес и как он используется в машинном обучении?
18. В чём заключаются преимущества и недостатки случайного леса по сравнению с другими методами машинного обучения?
19. Что такое базовые нейросетевые подходы и как они работают?
20. Что такое метод обратного распространения ошибки и как он используется для обучения нейросетей?
21. Что такое проблема затухания градиента и как она решается в машинном обучении?
22. Что такое представление текстовых видов и как оно используется в машинном обучении?
23. Что такое предобработка текстов и как она влияет на качество моделей машинного обучения?
24. Что такое рекуррентные нейронные сети и как они работают?
25. Что такое метод обратного распространения ошибки по времени и как он используется в рекуррентных нейронных сетях?
26. Что такое слои LSTM и GRU и как они используются в рекуррентных нейронных сетях?
27. Что такое свёрточные нейронные сети и как они работают?
28. Что такое слои MaxPooling и average pooling и как они используются в свёрточных нейронных сетях?
29. Что такое концепция «головы» и «хвоста» в свёрточной нейронной сети и как она применяется?
30. Что такое обучение без учителя и какие задачи оно решает?
31. Что такое метрики качества для обучения без учителя и как они работают?
32. Что такое методы кластеризации и как они используются в обучении без учителя?
33. Что такое методы визуализации и как они применяются для анализа данных?

Примеры экзаменационных вопросов в программе

Билет 1.

1. Что такое пороговая функция потерь и как она используется в задачах классификации?
2. Что такое ансамблевые подходы и как они улучшают точность моделей машинного обучения?

Билет 2.

1. Что такое метод обратного распространения ошибки и как он используется для обучения нейросетей?
2. Что такое предобработка текстов и как она влияет на качество моделей машинного обучения?

Билет 3.

1. В чём заключается подход Bootstrap Aggregation и как он работает?
2. Что такое слои MaxPooling и average pooling и как они используются в свёрточных нейронных сетях?

Критерии оценивания

- оценка «отлично (10)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений
- оценка «хорошо (7)» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (6)» выставляется студенту, если он знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (5)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «удовлетворительно (4)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Экзамен может проводиться по итогам текущей успеваемости и сдачи заданий и других видов работ, предусмотренных программой дисциплины и (или) путем организации специального опроса, проводимого в устной и (или) письменной форме.

При проведении экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося не должен превышать одного астрономического часа.

Во время проведения экзамена обучающиеся могут пользоваться программой дисциплины, а также справочной литературой, конспектами лекций или другими материалами.