



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ГЕОДЕЗИИ И КАРТОГРАФИИ
(МИИГАиК)

Факультет геоинформатики и информационной безопасности
Кафедра информатики и геоинформационных технологий



**ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ**

Сертификат: 5387FD2AF5B17FED816FDE9A1D4973B1
Владелец: Степанченко Алексей Леонидович
Действителен: с 21.07.2025 по 14.10.2026

«УТВЕРЖДАЮ»

Проректор
А. Л. Степанченко
16 февраля 2024 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
ОСНОВЫ КОМПЬЮТЕРНЫХ НАУК
«Б1.О.05.05 ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ»**

Уровень высшего образования: бакалавриат

Направление подготовки: 09.03.03 Прикладная информатика

Направленность(профиль) подготовки: Прикладная информатика

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Год набора (приема на обучение): 2024

Срок получения образования: 4 года

Объем:
в зачетных единицах: 4 з.е.
в академических часах: 144 ак.ч.

Москва, 2024

Разработчики:

Старший преподаватель кафедры Информатики и геоинформационных технологий Стрельцов А. С.

Рабочая программа дисциплины (модуля) составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.03.03 Прикладная информатика, утвержденного приказом Минобрнауки от 19.09.2017 № 922

Согласование и утверждение

№	Подразделение	Ответственное лицо	ФИО	Виза	Дата, протокол (при наличии)
1	факультет Геоинформатики и информационной безопасности	Председатель ученого совета факультета	Матерухин А. В.	Рассмотрено	14.12.2023, № 7-23

Содержание

1. Цель и задачи освоения дисциплины (модуля)
2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы
3. Место дисциплины в структуре ОП
4. Объем дисциплины и виды учебной работы
5. Содержание дисциплины
 - 5.1. Разделы, темы дисциплины и виды занятий
 - 5.2. Содержание разделов, тем дисциплины и формы текущего контроля
6. Оценочные материалы текущего контроля
7. Оценочные материалы промежуточной аттестации
8. Порядок проведения промежуточной аттестации
9. Материально-техническое и учебно-методическое обеспечение дисциплины
 - 9.1. Перечень основной и дополнительной учебной литературы
 - 9.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся
 - 9.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине
 - 9.4. Оборудование и технические средства обучения
10. Методические указания по освоению дисциплины (модуля)

1. Цель и задачи освоения дисциплины (модуля)

Цель освоения дисциплины - формирование у обучающихся фундаментальных знаний в области машинного обучения и практических навыков разработки, обучения и развертывания моделей машинного обучения с использованием современных библиотек Python для решения задач анализа данных, прогнозирования и классификации в различных предметных областях.

Задачи изучения дисциплины:

- Освоение математических и программных основ анализа данных, включая работу с библиотеками NumPy для численных вычислений, Pandas для обработки и анализа структурированных данных, Matplotlib и Plotly для визуализации данных, а также изучение основных метрик оценки качества моделей в задачах регрессии и классификации.;
- Изучение и практическое применение классических алгоритмов машинного обучения с учителем, включая метод k-ближайших соседей (KNN), линейную и логистическую регрессию, решающие деревья, случайный лес, метод наивного Байеса, метод опорных векторов, а также освоение методов подбора оптимальных гиперпараметров и техник кросс-валидации.;
- Формирование навыков построения комплексных решений на основе ансамблевых методов машинного обучения, разработки систем классификации текстов и других типов данных, а также компетенций в области развертывания обученных ML-моделей в продуктивной среде с использованием современных инструментов и фреймворков..

2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы

Компетенции, индикаторы и результаты обучения

ОПК-1 Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности

ОПК-1.1 Разрабатывает и валидирует математические модели, основанные на естественнонаучных и общетехнических принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

Знать:

ОПК-1.1/Зн1 Знание основных математических и физических законов и моделей, основанных на естественнонаучных и общетехнических принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

Уметь:

ОПК-1.1/Ум1 умение применять математические и физические законы и модели, основанные на естественнонаучных и общетехнических принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

ОПК-1.1/Ум2 Решать стандартные профессиональные задачи с применением естественнонаучных и общетехнических знаний, методов математического анализа и моделирования

Владеть:

ОПК-1.1/Нв1 владение навыками применения основных математических и физических законов и моделей, основанных на естественнонаучных и общетехнических принципах, для описания, прогнозирования и оптимизации процессов и систем в профессиональной сфере, а также интерпретирует полученные результаты

ОПК-1.2 Интегрирует естественнонаучные и общетехнические знания с методами математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решений

Знать:

ОПК-1.2/Зн1 знание методов математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решений

Уметь:

ОПК-1.2/Ум1 умение применить методы математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решений

Владеть:

ОПК-1.2/Нв1 владение навыками применения методов математического анализа и моделирования для системного анализа профессиональных задач и разработки обоснованных решений

ОПК-1.3 Планирует, проводит и интерпретирует теоретические и экспериментальные исследования, выбирая адекватные методы, собирая, обрабатывая и анализируя данные, и критически оценивает их достоверность

Знать:

ОПК-1.3/Зн1 Знание физических законов и требований для планирования, проведения и интерпретации теоретических и экспериментальных исследований, выбора адекватных методов, обработки и анализа данных, и критической оценки их достоверности

Уметь:

ОПК-1.3/Ум1 умение применить физические законы и требования для планирования, проведения и интерпретации теоретических и экспериментальных исследований, выбора адекватных методов, обработки и анализа данных, и критической оценки их достоверности

Владеть:

ОПК-1.3/Нв1 владение навыками применения физических законов для планирования, проведения и интерпретации теоретических и экспериментальных исследований, выбора адекватных методов, обработки и анализа данных, и критической оценки их достоверности

ОПК-1.4 Использует современные цифровые инструменты для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

Знать:

ОПК-1.4/Зн1 знание современных цифровых инструментов для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

Уметь:

ОПК-1.4/Ум1 умение применения современных цифровых инструментов для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

Владеть:

ОПК-1.4/Нв1 владение навыками применения современных цифровых инструментов для математических расчетов, моделирования и обработки исследовательских данных, а также для представления научно обоснованных решений

3. Место дисциплины в структуре ОП

Дисциплина (модуль) Б1.О.05.05 «Основы машинного обучения» относится к обязательной части образовательной программы и изучается в семестре(ах): 4.

В процессе изучения дисциплины студент готовится к решению типов задач профессиональной деятельности, предусмотренных ФГОС ВО и образовательной программой.

4. Объем дисциплины (модуля) и виды учебной работы

Период обучения	Общая трудоемкость (часы)	Общая трудоемкость (ЗЕТ)	Лекционные занятия (часы)	Практические занятия (часы)	Самостоятельная работа (часы)	Промежуточная аттестация (часы)
4 семестр	144	4	36	36	72	Зачет с оценкой
Всего	144	4	36	36	72	

5. Содержание дисциплины (модуля)

5.1. Разделы, темы дисциплины и виды занятий

(часы промежуточной аттестации не указываются)

Наименование раздела, темы	Всего	Лекционные занятия	Практические занятия	Самостоятельная работа	Планируемые результаты с обучения, соотношенные с результатами освоения программы
Раздел 1. Математические основы анализа данных	8	2	2	4	ОПК-1.3 ОПК-1.4
Тема 1.1. Библиотека NumPy для численных вычислений	4	1	1	2	
Тема 1.2. Статистические вычисления в NumPy	4	1	1	2	
Раздел 2. Визуализация и анализ данных	16	4	4	8	ОПК-1.3 ОПК-1.4
Тема 2.1. Визуализация данных с Matplotlib	4	1	1	2	
Тема 2.2. Интерактивная визуализация с Plotly	4	1	1	2	
Тема 2.3. Работа с табличными данными в Pandas	4	1	1	2	
Тема 2.4. Предобработка данных с Pandas	4	1	1	2	

Раздел 3. Основы теории машинного обучения	12	3	3	6	ОПК-1.3 ОПК-1.4
Тема 3.1. Введение в машинное обучение	4	1	1	2	
Тема 3.2. Метрики оценки качества моделей	4	1	1	2	
Тема 3.3. Валидация моделей и кросс-валидация	4	1	1	2	
Раздел 4. Методы обучения на основе расстояний	8	2	2	4	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 4.1. Метод k-ближайших соседей (KNN)	4	1	1	2	
Тема 4.2. Оптимизация гиперпараметров	4	1	1	2	
Раздел 5. Линейные модели	16	4	4	8	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 5.1. Линейная регрессия	8	2	2	4	
Тема 5.2. Логистическая регрессия	8	2	2	4	
Раздел 6. Древовидные модели	20	5	5	10	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 6.1. Решающие деревья для классификации	8	2	2	4	
Тема 6.2. Решающие деревья для регрессии	8	2	2	4	
Тема 6.3. Критерии останова и регуляризация деревьев	4	1	1	2	
Раздел 7. Ансамблевые методы	24	6	6	12	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 7.1. Случайный лес (Random Forest)	8	2	2	4	
Тема 7.2. Градиентный бустинг	8	2	2	4	
Тема 7.3. Стекинг и блендинг моделей	8	2	2	4	
Раздел 8. Вероятностные и дистанционные методы	16	4	4	8	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 8.1. Наивный байесовский классификатор	8	2	2	4	
Тема 8.2. Метод опорных векторов (SVM)	8	2	2	4	
Раздел 9. Практическое применение машинного обучения	24	6	6	12	ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4
Тема 9.1. Классификация текстовых данных	8	2	2	4	
Тема 9.2. Построение комплексных ML-решений	8	2	2	4	
Тема 9.3. Развертывание ML-моделей	8	2	2	4	
Итого	144	36	36	72	

5.2. Содержание разделов, тем дисциплин и формы текущего контроля

Раздел 1. Математические основы анализа данных

Тема 1.1. Библиотека NumPy для численных вычислений

Основы работы с многомерными массивами ndarray. Создание массивов различными способами. Индексация и срезы. Математические операции над массивами: поэлементные операции, матричные операции, векторизация вычислений. Функции агрегации. Broadcasting - правила транслирования. Линейная алгебра в NumPy.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 1.2. Статистические вычисления в NumPy

Генерация случайных чисел и распределений . Описательная статистика: среднее, медиана, дисперсия, стандартное отклонение. Корреляционный анализ. Работа с пропущенными значениями. Применение NumPy для предобработки данных в машинном обучении.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 2. Визуализация и анализ данных

Тема 2.1. Визуализация данных с Matplotlib

Основы построения графиков с библиотекой Matplotlib . Типы визуализаций: линейные графики, гистограммы, диаграммы рассеяния, box-plot. Настройка осей, легенд, заголовков. Создание подграфиков. Визуализация распределений и зависимостей в данных для машинного обучения.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 2.2. Интерактивная визуализация с Plotly

Создание интерактивных графиков с библиотекой Plotly . 3D-визуализации. Анимированные графики. Dashboard-визуализации. Визуализация результатов кластеризации и классификации. Построение интерактивных отчетов по результатам анализа.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 2.3. Работа с табличными данными в Pandas

Структуры данных Series и DataFrame . Загрузка данных из различных источников (CSV, Excel, SQL). Индексация и выборка данных. Группировка и агрегация. Объединение датафреймов. Работа с временными рядами. Pivot-таблицы.

Текущий контроль

Вид (форма) контроля

Просмотр выполненных работ

Тема 2.4. Предобработка данных с Pandas

Обработка пропущенных значений . Удаление дубликатов. Преобразование типов данных. Кодирование категориальных признаков. Нормализация и стандартизация данных. Feature engineering - создание новых признаков. Подготовка данных для машинного обучения.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 3. Основы теории машинного обучения

Тема 3.1. Введение в машинное обучение

Основные понятия и терминология машинного обучения. Типы задач: классификация, регрессия, кластеризация. Обучение с учителем и без учителя. Этапы разработки ML-модели. Проблемы переобучения и недообучения. Bias-variance tradeoff.

Тема 3.2. Метрики оценки качества моделей

Метрики для задач регрессии: MSE, RMSE, MAE, R^2 . Метрики для задач классификации: accuracy, precision, recall, F1-score. Матрица ошибок. ROC-кривая и AUC. Выбор оптимальной метрики в зависимости от бизнес-задачи. Интерпретация метрик качества.

Тема 3.3. Валидация моделей и кросс-валидация

Разделение данных на обучающую, валидационную и тестовую выборки. Методы кросс-валидации: k-fold, stratified k-fold, leave-one-out. Временная валидация для временных рядов. Оценка стабильности модели. Выбор оптимальной модели по результатам валидации.

Раздел 4. Методы обучения на основе расстояний

Тема 4.1. Метод k-ближайших соседей (KNN)

Алгоритм KNN для задач классификации и регрессии . Выбор метрики расстояния: евклидово, манхэттенское, расстояние Минковского. Влияние параметра k на качество модели. Взвешенный KNN. Масштабирование признаков. Применение KNN для рекомендательных систем.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 4.2. Оптимизация гиперпараметров

Методы подбора оптимальных гиперпараметров . Grid Search и Random Search. Байесовская оптимизация. Автоматизированный подбор гиперпараметров с помощью GridSearchCV и RandomizedSearchCV. Стратегии ускорения поиска. Визуализация результатов оптимизации.

Раздел 5. Линейные модели

Тема 5.1. Линейная регрессия

Метод наименьших квадратов . Множественная линейная регрессия. Полиномиальная регрессия. Регуляризация: Ridge, Lasso, ElasticNet. Градиентный спуск и его модификации. Диагностика модели: анализ остатков, проверка предположений.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 5.2. Логистическая регрессия

Бинарная и многоклассовая логистическая регрессия . Функция логистических потерь. Интерпретация коэффициентов. Регуляризация в логистической регрессии. Калибровка вероятностей. Применение для задач кредитного скоринга и медицинской диагностики.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 6. Древоподобные модели

Тема 6.1. Решающие деревья для классификации

Алгоритм построения дерева решений . Критерии разбиения: Gini, энтропия. Обрезка дерева для борьбы с переобучением. Визуализация и интерпретация деревьев. Преимущества и недостатки деревьев решений.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 6.2. Решающие деревья для регрессии

Особенности деревьев регрессии . Критерии разбиения: MSE, MAE. Экстраполяционные свойства деревьев. Сравнение с линейными моделями. Применение для задач прогнозирования.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 6.3. Критерии останова и регуляризация деревьев

Параметры контроля сложности дерева . Минимальное количество объектов в листе. Максимальная глубина дерева. Минимальное улучшение критерия. Постпрuning и препруning. Влияние параметров на качество модели.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 7. Ансамблевые методы

Тема 7.1. Случайный лес (Random Forest)

Принцип работы Random Forest . Бэггинг и случайные подпространства признаков. Out-of-bag оценка. Важность признаков. Параметры настройки: количество деревьев, глубина, количество признаков. Применение для задач с большим числом признаков.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 7.2. Градиентный бустинг

Основы бустинга . Алгоритм градиентного бустинга. XGBoost, LightGBM, CatBoost - сравнение реализаций. Настройка гиперпараметров бустинга. Early stopping. Обработка категориальных признаков.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 7.3. Стекинг и блендинг моделей

Методы комбинирования моделей . Усреднение предсказаний. Взвешенное голосование. Стекинг с мета-алгоритмом. Блендинг на валидационной выборке. Создание ансамблей для соревнований по машинному обучению.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 8. Вероятностные и дистанционные методы

Тема 8.1. Наивный байесовский классификатор

Теорема Байеса и ее применение . Гауссовский наивный Байес. Мультиномиальный и бернуллиевский варианты. Сглаживание Лапласа. Применение для классификации текстов и фильтрации спама.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 8.2. Метод опорных векторов (SVM)

Линейный SVM для разделимых и неразделимых выборок . Kernel trick и нелинейные SVM. Типы ядер: RBF, полиномиальное, сигмоидальное. Параметры C и gamma. SVM для регрессии (SVR). Многоклассовая классификация.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Раздел 9. Практическое применение машинного обучения

Тема 9.1. Классификация текстовых данных

Предобработка текста: токенизация, лемматизация, удаление стоп-слов. Векторизация текста: Bag of Words, TF-IDF, Word2Vec. Применение различных алгоритмов для классификации текстов. Анализ тональности. Классификация документов.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

Тема 9.2. Построение комплексных ML-решений

Создание end-to-end pipeline для машинного обучения. Автоматизация предобработки данных. Feature engineering и отбор признаков.

Тема 9.3. Развертывание ML-моделей

Сериализация обученных моделей с pickle и joblib. Создание REST API для моделей с Flask/FastAPI. Развертывание в облаке. Разработка графического интерфейса в Streamlit.

Текущий контроль

Вид (форма) контроля
Просмотр выполненных работ

6. Порядок проведения промежуточной аттестации

Порядок проведения промежуточной аттестации регламентируется локальными нормативными актами Университета:

- Положением о формах, периодичности, порядке текущего контроля успеваемости и промежуточной аттестации обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.6 «Организация и проведение промежуточной аттестации».
- Положением о порядке осуществления образовательного процесса, проведения текущего контроля успеваемости и промежуточной аттестации обучающихся с применением электронного обучения и дистанционных образовательных технологий в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный университет геодезии и картографии», п.8 «Проведение промежуточной аттестации с применением ЭО и ДОТ».

Промежуточная аттестация: Зачет с оценкой, 4 семестр.

Дифференцированный зачет по дисциплине проводится в формате письменного тестирования для итоговой оценки знаний.

Студентам будет предложен тест, состоящий из тридцати вопросов закрытого типа с несколькими вариантами ответов. На выполнение всей работы отводится шестьдесят минут, в течение которых запрещается использование любых вспомогательных материалов и электронных устройств.

Итоговая оценка выставляется в зависимости от процента правильных ответов. Для получения оценки «отлично» необходимо верно ответить на 90-100% вопросов (27-30

правильных ответов). Оценка «хорошо» ставится за 70-89% правильных ответов (21-26), а «удовлетворительно» — за 55-69% (17-20). Результат ниже 55% правильных ответов признается неудовлетворительным.

7. Оценочные материалы текущего контроля

Раздел 1. Математические основы анализа данных

Контролируемые ИДК: ОПК-1.3 ОПК-1.4

Тема 1.1. Библиотека NumPy для численных вычислений

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 1.2. Статистические вычисления в NumPy

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 2. Визуализация и анализ данных

Контролируемые ИДК: ОПК-1.3 ОПК-1.4

Тема 2.1. Визуализация данных с Matplotlib

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 2.2. Интерактивная визуализация с Plotly

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 2.3. Работа с табличными данными в Pandas

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 2.4. Предобработка данных с Pandas

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 4. Методы обучения на основе расстояний

Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 4.1. Метод k -ближайших соседей (KNN)

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 5. Линейные модели

Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 5.1. Линейная регрессия

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 5.2. Логистическая регрессия

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 6. Древовидные модели

Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 6.1. Решающие деревья для классификации

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 6.2. Решающие деревья для регрессии

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 6.3. Критерии останова и регуляризация деревьев

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 7. Ансамблевые методы

Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 7.1. Случайный лес (Random Forest)

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 7.2. Градиентный бустинг

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 7.3. Стекинг и блендинг моделей

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 8. Вероятностные и дистанционные методы

Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 8.1. Наивный байесовский классификатор

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 8.2. Метод опорных векторов (SVM)

Форма контроля/оценочное средство: Просмотр выполненных работ

Раздел 9. Практическое применение машинного обучения

Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Тема 9.1. Классификация текстовых данных

Форма контроля/оценочное средство: Просмотр выполненных работ

Тема 9.3. Развертывание ML-моделей

Форма контроля/оценочное средство: Просмотр выполненных работ

8. Оценочные материалы промежуточной аттестации

4 семестр, Зачет с оценкой

Контролируемые ИДК: ОПК-1.1 ОПК-1.2 ОПК-1.3 ОПК-1.4

Вопросы/Задания:

1. Что является основной структурой данных в библиотеке NumPy?

- A. DataFrame
- B. Series
- C. ndarray
- D. List

2. Как в NumPy называется набор правил для выполнения математических операций над массивами разной формы?

- A. Векторизация (Vectarization)

- B. Транслирование (Broadcasting)
- C. Агрегация (Aggregation)
- D. Индексация (Indexing)

3. Какие две основные структуры данных предоставляет библиотека Pandas?

- A. ndarray и Matrix
- B. Series и DataFrame
- C. List и Dictionary
- D. JSON и CSV

4. Как называется процесс создания новых признаков из существующих данных при работе с Pandas?

- A. Агрегация данных
- B. Инжиниринг признаков (Feature engineering)
- C. Нормализация данных
- D. Создание Pivot-таблиц

5. Какой тип визуализации НЕ упоминается в качестве базового для библиотеки Matplotlib в предоставленном содержании?

- A. Гистограммы
- B. Диаграммы рассеяния
- C. 3D-визуализации
- D. Box-plot

6. Что из перечисленного НЕ является основной задачей машинного обучения?

- A. Классификация
- B. Регрессия
- C. Визуализация
- D. Кластеризация

7. Как называется проблема, когда модель отлично работает на обучающих данных, но плохо справляется с новыми, неизвестными данными?

- A. Недообучение
- B. Переобучение
- C. Смещение (Bias)
- D. Дисперсия (Variance)

8. Какая из этих метрик обычно используется для оценки качества моделей регрессии?

- A. Accuracy
- B. Precision
- C. F1-score
- D. MSE (Mean Squared Error)

9. Какая пара метрик является ключевой для оценки моделей классификации и часто используется при построении матрицы ошибок?

- A. MSE и MAE
- B. Precision и Recall
- C. R^2 и RMSE
- D. Gini и Entropy

10. Какова основная цель k-fold кросс-валидации?

- A. Ускорить обучение модели
- B. Получить более стабильную и надежную оценку качества модели
- C. Визуализировать данные
- D. Уменьшить количество признаков

11. Что означает параметр 'k' в алгоритме k-ближайших соседей (KNN)?

- A. Количество используемых признаков
- B. Количество итоговых кластеров
- C. Количество соседних точек, учитываемых при предсказании
- D. Количество эпох обучения

12. Какой метод оптимизации гиперпараметров перебирает все возможные комбинации из заданного набора значений?

- A. Random Search
- B. Grid Search
- C. Байесовская оптимизация
- D. Градиентный спуск

13. Какой тип регуляризации в линейной регрессии способен полностью обнулять коэффициенты у наименее важных признаков?

- A. Ridge (L2)
- B. Lasso (L1)
- C. ElasticNet
- D. Полиномиальная

14. Какой из перечисленных критериев используется для разделения узлов в дереве решений для задачи классификации?

- A. MSE (Среднеквадратичная ошибка)
- B. MAE (Средняя абсолютная ошибка)
- C. Неопределенность Джини (Gini)
- D. Коэффициент детерминации R^2

15. Как называется процесс удаления ветвей из дерева решений для борьбы с переобучением?

- A. Разделение (Splitting)
- B. Рост (Growing)
- C. Обрезка (Pruning)
- D. Агрегация (Aggregation)

16. В чем заключается основной принцип работы ансамблевого метода "Случайный лес" (Random Forest)?

- A. Последовательное построение моделей, исправляющих ошибки предыдущих
- B. Объединение множества деревьев решений, обученных на разных подвыборках данных
- C. Поиск оптимальной гиперплоскости, разделяющей классы
- D. Использование теоремы Байеса для вычисления вероятностей

17. Что из перечисленного является популярной реализацией алгоритма градиентного бустинга?

- A. KNN
- B. SVM
- C. XGBoost
- D. NumPy

18. Какую роль выполняет "мета-алгоритм" в методе стекинга?

- A. Осуществляет предобработку исходных данных
- B. Обучает базовые модели ансамбля
- C. Объединяет предсказания базовых моделей для формирования итогового ответа
- D. Отбирает лучшие признаки для обучения

19. Для какой задачи, согласно содержанию, особенно хорошо подходит наивный байесовский классификатор?

- A. Распознавание изображений
- B. Прогнозирование временных рядов
- C. Фильтрация спама и классификация текстов
- D. Решение задач регрессии

20. В чем заключается "трюк с ядром" (kernel trick) в методе опорных векторов (SVM)?

- A. В ускорении процесса обучения
- B. В работе с нелинейно разделимыми данными путем их отображения в пространство более высокой размерности
- C. В уменьшении количества опорных векторов
- D. В автоматическом подборе параметра регуляризации

21. Как расшифровывается аббревиатура TF-IDF, используемая при векторизации текстов?

- A. Text Frequency - Inverse Data Frequency
- B. Term Frequency - Inverse Document Frequency
- C. Total Frequency - Inverted Data Format
- D. Text Format - Inverse Document Format

22. Какая библиотека упоминается для сериализации (сохранения) обученных моделей машинного обучения?

- A. NumPy
- B. Pandas
- C. Pickle
- D. Matplotlib

23. Какая технология упоминается для создания контейнеров, позволяющих легко развертывать ML-модели?

- A. Flask
- B. FastAPI
- C. Docker
- D. Git

24. Какая ключевая особенность библиотеки Plotly по сравнению с Matplotlib отмечена в программе?

- A. Она работает быстрее для простых статических графиков
- B. Она в основном используется для статистических вычислений
- C. Она специализируется на создании интерактивных и анимированных визуализаций
- D. Она является частью стандартной библиотеки Python

25. Какой алгоритм оптимизации упоминается как фундаментальный для обучения линейных моделей?

- A. K-Means
- B. Градиентный спуск
- C. Алгоритм построения деревьев
- D. Случайный лес

26. В чем заключается основная идея "бустинга"?

- A. В параллельном обучении множества независимых моделей
- B. В последовательном построении моделей, где каждая новая модель исправляет ошибки предыдущей
- C. В усреднении предсказаний моделей разных типов
- D. В выборе случайного подмножества признаков для каждой модели

27. Как называется процесс приведения слов к их базовой или словарной форме?

- A. Токенизация
- B. Удаление стоп-слов
- C. Лемматизация
- D. Векторизация

28. Компромисс между способностью модели минимизировать ошибку на обучающем наборе и ее способностью обобщаться на новые данные известен как:

- A. Компромисс точности и полноты (Accuracy-Precision tradeoff)
- B. Компромисс смещения и дисперсии (Bias-variance tradeoff)
- C. Компромисс k-fold
- D. Компромисс Ridge и Lasso

29. Что из перечисленного является типом ядра, используемого в методе опорных векторов (SVM)?

- A. Евклидово
- B. Манхэттенское
- C. RBF (Радиальная базисная функция)
- D. Косинусное

30. Какова цель создания REST API для модели машинного обучения?

- A. Повысить точность модели
- B. Сделать предсказания модели доступными для других приложений по сети
- C. Выполнить кросс-валидацию
- D. Визуализировать производительность модели

9. Материально-техническое и учебно-методическое обеспечение дисциплины

9.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Митяков Е. С. Искусственный интеллект и машинное обучение: учебное пособие для вузов: учебное пособие для вузов / Е. С. Митяков, А. Г. Шмелева, А. И. Ладынин. - Санкт-Петербург: Лань, 2025. - 252 с. - 978-5-507-51465-6. - Текст: электронный. // Издательство Лань: [сайт]. - URL: <https://e.lanbook.com/img/cover/book/450827.jpg> (дата обращения: 16.03.2025). - Режим доступа: по подписке

2. Титов А. Н. Обработка данных в Python. Основы работы с библиотекой Pandas: учебно-методическое пособие: учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. - Обработка данных в Python. Основы работы с библиотекой Pandas - Казань: Издательство КНИТУ, 2022. - 116 с. - 978-5-7882-3164-8. - Текст: электронный. // IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/129244.html> (дата обращения: 16.03.2025). - Режим доступа: по подписке

Дополнительная литература

1. Пальмов С. В. Интеллектуальный анализ данных: учебное пособие: учебное пособие / С. В. Пальмов. - Интеллектуальный анализ данных - Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. - 127 с. - 2227-8397. - Текст: электронный. // IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/75376.html> (дата обращения: 16.03.2025). - Режим доступа: по подписке

9.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся

Профессиональные базы данных

Не используются.

Ресурсы «Интернет»

1. <https://docs.python.org/3/index.html> - Техническая документация по языку программирования Python
2. <https://sdo.miigaik.ru/> - Система дистанционного обучения МИИГАиК

9.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения

(обновление производится по мере появления новых версий программы)

1. Anaconda;
2. Microsoft Visual Studio;

Перечень информационно-справочных систем

(обновление выполняется еженедельно)

1. 1С Предприятие;

9.4. Оборудование и технические средства обучения

Учебная аудитория

Маркерная доска - 1 шт.

Моноблок - 96 шт.

Проектор - 8 шт.

Автоматизированное рабочее место (АРМ) - 40 шт.

Плоттер - 1 шт.

Сканер - 1 шт.

Интерактивная панель - 1 шт.

10. Методические указания по освоению дисциплины (модуля)

С целью успешного освоения данного предмета студенту необходимо выполнять домашние работы по указанным преподавателям темам и представленным на занятиях материалам. Для более успешного выполнения работы студенту предлагается использование учебно-методических материалов, соответствующих указанной теме.

Самостоятельная работа студентов описывается и регулируется методическими рекомендациями по дисциплине.

Описание возможностей изучения дисциплины лицами с ОВЗ и инвалидами

При необходимости обучения студентов-инвалидов и лиц с ограниченными возможностями здоровья аудиторные занятия могут быть заменены или дополнены изучением полнотекстовых лекций, презентаций, видео- и аудиоматериалов в электронной информационно-образовательной среде (ЭИОС) университета. Индивидуальные задания подбираются в адаптированных к ограничениям здоровья формах (письменно или устно, в форме презентаций). Выбор методов обучения зависит от их доступности для инвалидов и лиц с ограниченными возможностями здоровья.

Рабочая программа дисциплины для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.