МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Методические указания к практическим работам

Направление подготовки:

09.04.02 «Информационные системы и технологии»

Квалификация выпускника: магистр

Ставрополь, 2025

Н 63 **Системы искусственного интеллекта:** Методические указания к практическим работам для студентов направления 09.04.02 «Информационные системы и технологии» / Николаев Е.И. – Ставрополь: Изд-во СКФУ, 2025. – 30 с.

Учебное пособие (практикум) по дисциплине «Системы искусственного интеллекта» для студентов направления 09.04.02 «Информационные системы и технологии». Пособие охватывает практические аспекты построения информационных систем на основе методов искусственного интеллекта с применением современных технологий разработки информационных систем и инструментария анализа данных. Основное внимание уделяется теории обучения машин (машинное обучение, machine learning).

Автор:

канд. техн. наук, доцент **Е.И. Николаев**

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Методические указания к практическим работам по дисциплине «Системы искусственного интеллекта» для студентов направления 09.04.02 «Информационные системы и технологии». Пособие охватывает теоретические аспекты построения информационных систем на основе методов искусственного интеллекта. Основное внимание уделяется теории обучения машин (машинное обучение, machine learning).

Многие задачи, возникающие в практических приложениях, не могут быть решены заранее известными методами или алгоритмами. Это происходит по той причине, что нам заранее не известны механизмы порождения исходных данных или же известная нам информация недостаточна для построения модели источника, генерирующего поступающие к нам данные. Как говорят, мы получаем данные из «черного ящика». В этих условиях ничего не остается, как только изучать доступную нам последовательность исходных данных и пытаться строить предсказания совершенствуя нашу схему в процессе предсказания. Подход, при котором про- шлые данные или примеры используются для первоначального формирования и совершенствования схемы предсказания, называется методом машинного обучения (Machine Learning). Машинное обучение – чрезвычайно широкая и динамически развивающаяся область исследований, использующая огромное число теоретических и практических методов. Данное пособие ни в какой мере не претендует на какое-либо исчерпывающее изложение содержания данной области. Основная цель – дать студентам теоретическое представление о современных математических проблемамах в области систем искусственного интеллекта, а также познакомить с путями их решения.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 1. ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ. СРЕДСТВА МАШИННОГО ОБУЧЕНИЯ

1. Цели и задачи

Цель лабораторной работы: изучение программных средств для организации рабочего места специалиста по анализу данных и машинному обучению.

Основные задачи:

– установка и настройка среды разработки Python;

– изучение принципов загрузки и очистки данных;

– получение навыков по предварительной обработке данных на языке Python;

– изучение основных библиотек Python для работы с данными.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с базовыми принципами языка Python, используя следующие источники: [1-4].

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Выбор набора данных.
2. Основные характеристики наборов данных, методы их получения.
3. Признаки. Типология признаков.
4. Библиотеки анализа данных.
5. Языки и стеки технологий для анализа данных.
6. Библиотека Pandas

7. Библиотека numpy.

8. Библиотека Matplotlib.

9. Библиотека Seaborn.

10. Графическое представление наборов данных: основные типы визуализаций, их назначение.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-6].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 2. СРЕДСТВА ЯЗЫКА PYTHON

1. Цели и задачи

Цель лабораторной работы: изучение программных средств для организации рабочего места специалиста по анализу данных и машинному обучению.

Основные задачи:

– проведение первичного анализа набора данных;

– изучение принципов выбора метода машинного обучения;

– получение навыков по предварительной обработке данных на языке Python;

– изучение основных библиотек Python для работы с данными.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с базовыми принципами языка Python, используя следующие источники: [1-4].

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Построение графиков с использованием библиотеки matplotlib.

2. Загрузка данных из различных источников.

3. Обработка пропущенных значений. Различные стратегии обработки пропусков в данных.

4. Обработка категориальных признаков.

5. Кодировка категориальных признаков. Проблема мультиколлениарности при использовании фиктивных признаков.

6. Целевые признаки и методы их обработки.

7. Средстваа Python для работы с изображениями.

8. Сервисные функции класса DataFrame.

9. Статистические параметры набора данных и методы их получения.

10. Визуализация многомерных данных.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-6].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 3. ВИЗУАЛИЗАЦИЯ ДАННЫХ

1. Цели и задачи

Цель лабораторной работы: изучение программных средств для визуализации наборов данных.

Основные задачи:

– установка и настройка matplotlib, seaborn;

–изучение основных типов графиков библиотеки matplotlib;

–изучение основных типов графиков библиотеки seaborn;

– получение навыков анализа данных по визуальным представлениям данных.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с базовыми принципами языка Python, используя следующие источники: [1-4]. Особое внимание необходимо уделить репозитарию с исходными кодами.

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Функция DataFrame.info( ).
2. Функция DataFrame.hist( ).
3. Построение диаграмм с использованием pyplot.bar( ). Задание параметров диаграммы.
4. Методы задания подписей к осям. Метки наборов данных.
5. График «ящик с усами» (seaborn.boxplot).
6. Анализ нескольких признаков с использованием seaborn.boxplot.

7. Интерквартильный размах распределения, квартили, перцентили.

8. Отображение попарного распределения признаков.

9. График pyplot.scactter.

10. Построение корреляционных матриц. Значение элементов корреляционной матрицы. Визуализация корреляционной матрицы.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-4].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 4. АНАЛИЗ НАБОРА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ СРЕДСТВ ВИЗУАЛИЗАЦИИ

1. Цели и задачи

Цель лабораторной работы: изучение программных средств для визуализации наборов данных.

Основные задачи:

– установка и настройка matplotlib, seaborn;

–изучение основных типов графиков библиотеки matplotlib;

–изучение основных типов графиков библиотеки seaborn;

– получение навыков анализа данных по визуальным представлениям данных.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с базовыми принципами языка Python, используя следующие источники: [1-4]. Особое внимание необходимо уделить репозитарию с исходными кодами.

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Объектный подход к построению графиков.
2. Тепловые карты.
3. Использование фигур и нострайка осей в matplotlib.
4. Методы imshow( ) и imsave( ), savefig( ).
5. Средства визуализации Pandas.
6. Библиотека plotly.

7. Библиотека d3.js.

8. Построение нескольких графиков.

9. График pyplot.scactter.

10. Построение корреляционных матриц. Значение элементов корреляционной матрицы. Визуализация корреляционной матрицы.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-4].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 5. KNN-КЛАССИФИКАЦИЯ

1. Цели и задачи

Цель лабораторной работы: изучение принципов построения информационных систем с использованием метрических методов классификации.

Основные задачи:

– изучение инструментария Python для реализации алгоритмов метрической классификации;

– изучение методов оптимизации параметров метрической классификации;

– освоение модификаций kNN-метода.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с теорией построения метрических классификаторов, используя следующие источники: [1-4]. Особое внимание необходимо уделить репозитарию с исходными кодами.

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Метод ближайшего соседа (NN-метод).
2. Метод k ближайших соседей (kNN-метод).
3. Выбор оптимального количества соседей.
4. Обучение kNN-модели.
5. Тюнинг kNN-модели.

6. Различные математические постановки kNN-метода (метод функций энергии).

7. Преимущества, недостатки и области применения kNN-метода.

8. Способы задания функции расстояния.

9. Использование средств языка Python для обучения kNN-классификатора.

10. Использование средств языка Python для тюнинга kNN-классификатора и визуализации результатов.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-4].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 6. ЛОГИЧЕСКИЕ ДЕРЕВЬЯ

1. Цели и задачи

Цель лабораторной работы: изучение принципов построения информационных систем с использованием логических методов классификации.

Основные задачи:

– освоение технологии внедрения алгоритмов на онове решающих списков в приложения;

– освоение технологии внедрения алгоритмов на онове решающих деревьев в приложения;

– изучение параметров логической классификации;

– освоение модификаций логических методов классификации.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с теорией построения логических классификаторов, используя следующие источники: [1-4].

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Библиотека sklearn. Общая характеристика и возможности. Основные модули библиотеки.

2. Класс sklearn.tree.DecisionTreeClassifier. Параметры классификатора DecisionTreeClassifier.

3. Обучение модели DecisionTreeClassifier.

4. Тюнинг модели DecisionTreeClassifier.

5. Визуализация деревьев решений.

6. Модуль sklearn.model_selection.

7. Класс FridSearchCV.

8. Пайплайн модели машинного обучения.

9. Визуализация результатов классификации на основе дерева решений.

10. Ошибки классификации, критерии точности классификации.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-4].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 7. ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

1. Цели и задачи

Цель лабораторной работы: изучение принципов построения информационных систем с использованием линейных методов машинного обучения.

Основные задачи:

– освоение технологии внедрения алгоритмов линейной классификации в приложения;

– изучение основных приемов рабты с разреженными матрицами в ходе машинного обучения;

– освоение техники построения, обучения и оценки модели логистической регресии;

– освоение приемов работы с синтезированными признаками, масштабированием и настройкой гипперпараметров.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с теорией построения регрессионных моделей машинного обучения, используя следующие источники: [1-4].

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Модуль sklearn.preprocessing. Общая характеристика и возможности.
2. Нахождение агрегирующих показателей для набора данных.
3. Этапы разработки модели машинного обучения.

4. Линейные модели машинного обучения. Линейно разделимая выборка.

5. Масштабирование признаков. Масштабирование категориальных признаков.

6. Построение кривой AUC-ROC. Определение оптимальных параметров.

7. Основные параметры линеймой модели классификации. Гипперпараметры линейной модели.

8. Метод градиентного спуска.

9. Метод стахостического градиента.

10. математическая постановка задачи логистической регрессии.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-4].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 8. ЛИНЕЙНАЯ РЕГРЕССИЯ

1. Цели и задачи

Цель лабораторной работы: изучение принципов построения информационных систем с использованием линейных методов машинного обучения.

Основные задачи:

– освоение методологии работы с моделями линейной регресии в задачах машинного обучения;

– освоение метдик работы с линейными моделями в python;

– освоение методики применения методов регрессии;

– изучение основных параметров регрессионных моделей.

2. Теоретическая часть

Перед выполнением лабораторной работы необходимо ознакомиться с теорией построения регрессионных моделей машинного обучения, используя следующие источники: [1-4].

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Математическая постановка задачи регрессии. Типы решаемых задач.
2. Класс TfidfVectorizer. Назначение и параметры.
3. TF-IDF-матрица. Назначение и метод построения.
4. Класс Ridge.
5. Использование фиктивных признаков в линейной регрессии.

6. Построение кривой AUC-ROC. Определение оптимальных параметров.

7. Основные параметры линеймой модели классификации. Гипперпараметры линейной модели.

8. Метод градиентного спуска.

9. Метод стахостического градиента.

10. Вычисление MSE.

4. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1–4].

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 9. ПОЛИНОМИАЛЬНАЯ РЕГРЕССИЯ

1. Цели и задачи

Цель практической работы: научиться применять разработанный пайплайн для тиражирования кода с целью решения задачи полиномиальной регрессии.

Основные задачи:

– получение навыков рефакторинга кода в проектах машинного обучения;

– изучение поведения модели полиномиальной регрессии при изменении степени полинома.

2. Теоретическая часть

Линейная и параболическая модели являются частными случаями более сложной модели – полиномиальной. Построить модель регрессии – это значит найти параметры той функции, которая будет в ней фигурировать. Для линейной регрессии – два параметра: коэффициент и свободный член.

Полиномиальная регрессия может применяться в математической статистике при моделировании трендовых составляющих временных рядов. Временной ряд – это, по сути, ряд чисел, которые зависят от времени. Например, средние значения температуры воздуха по дням за прошедший год, или доход предприятия по месяцам. Порядок моделируемого полинома оценивается специальными методами, например, критерием серий. Цель построения модели полиномиальной регрессии в области временных рядов всё та же – прогнозирование.

3. Вопросы и задания

В рамках практического занятия необходимо рассмотреть следующие вопросы:

1. Математическая постановка задачи полиномиальной регрессии. Типы решаемых задач.

2. Возможности библиотеки Python для решения задач полиномиальной регрессии.

3. Применение масштабирования признаков при полиномиальной регрессии.

4. Категориальные признаки при полиномиальной регрессии.

5. Класс PolynomialFeatures.

6. Полиномиальная регрессия как линейная регрессия на полиномиальных признаках.

7. Трансформация признаков при использовании готовой полиномиальной модели регрессии.

8. Степень полиномиальной регрессии. Выбор степени полинома.

9. Визуализация результатов вычислений при моделировании процессов с использованием полиномиальной регрессии.

10. Вычисление MSE.

4. Контрольные вопросы

1. Почему при реализации многомерной линейной регрессии необходимо добавить фиктивный признак с единственным значением 1.0?.

2. то такое фиктивная переменная? Поясните причину удаления одной фиктивной переменной, возникающей при перекодировке категориального признака.

3. С использованием какого класса создается модель полиномиальной регрессии?

4. Поясните принцип преобразования признаков при построении полиномиальной регрессии.

5. Возможно ли применение технологий масштабирования признаков при реализации полиномиальной регрессии?

5. Список литературы

Для выполнения лабораторной работы, при подготовке к защите, а также для ответа на контрольные вопросы рекомендуется использовать следующие источники: [1-4].

ЗАКЛЮЧЕНИЕ

Учебное пособие (лабораторный практикум) по дисциплине «Системы искусственного интеллекта» для студентов направления 09.04.02 «Информационные системы и технологии». Пособие охватывает теоретические аспекты построения информационных систем на основе методов искусственного интеллекта, а также предлагает студентам практические рекомендации по разработке систем на основе мтеодов искусственного интеллекта. Основное внимание уделяется теории обучения машин (машинное обучение, machine learning).

В пособии рассмотрены практические аспекты проектирования и разработки информационных систем для решения различных задач с использованием следующих подходов: линейных методов классификации, аппарата нейронных сетей, байесовских методов классификации, алгоритмов восстановления регрессии, алгоритмов логической классификации.

Многие задачи, возникающие в практических приложениях, не могут быть решены заранее известными методами или алгоритмами. Это происходит по той причине, что нам заранее не известны механизмы порождения исходных данных или же известная нам информация недостаточна для построения модели источника, генерирующего поступающие к нам данные. Как говорят, мы получаем данные из «черного ящика». В этих условиях ничего не остается, как только изучать доступную нам последовательность исходных данных и пытаться строить предсказания совершенствуя нашу схему в процессе предсказания. Подход, при котором про- шлые данные или примеры используются для первоначального формирования и совершенствования схемы предсказания, называется методом машинного обучения (Machine Learning). Машинное обучение – чрезвычайно широкая и динамически развивающаяся область исследований, использующая огромное число теоретических и практических методов.

СПИСОК ЛИТЕРАТУРЫ

Список основной литературы

1. Уэс, Маккинли. Python и анализ данных Электронный ресурс / Маккинли Уэс ; пер. А. А. Слинкин. - Python и анализ данных,2025-04-19. - Саратов : Профобразование, 2017. - 482 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4488-0046-7, экземпляров неограниченно.

2. Сузи, Р.А. Язык программирования Python Электронный ресурс : учебное пособие / Р.А. Сузи. - Язык программирования Python,2020-07-28. - Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. - 350 с. - Книга находится в базовой версии ЭБС IPRbooks. - ISBN 5-9556-0058-2, экземпляров неограниченно

Список дополнительной литературы

3. Стенли, Липпман. Язык программирования C++ Электронный ресурс : Полное руководство / Липпман Стенли, Лажойе Жози ; пер. А. Слинкин. - Язык программирования C++,2025-04-19. - Саратов : Профобразование, 2017. - 1104 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4488-0136-5, экземпляров неограниченно

4. Седжвик, Р. Алгоритмы на C++ / Р. Седжвик. - 2-е изд., испр. - Москва : Национальный Открытый Университет «ИНТУИТ», 2016. - 1773 с., экземпляров неограниченно

Список интернет-ресурсов

5. https://archive.ics.uci.edu/ml/index.html – Репозиторий наборов данных для машинного обучения (Центр машинного обучения и интеллектуальных систем).

6. https://www.kaggle.com – Портал и система проведения соревнований по проблемам анализа данных.

Системы искусственного интеллекта

Методические указания к практическим работам

Автор

Николаев Евгений Иванович