Министерство образования и науки РФ

Федеральное государственное автономное

образовательное учреждение высшего образования

«Национальный исследовательский университет ИТМО»

**факультет программной инженерии и компьютерной техники**

**ЛАБОРАТОРНАЯ РАБОТА №4-7**

по дисциплине

‘Системы искусственного интеллекта‘

*Выполнил:*

Студент группы P33312

Соболев Иван Александрович

*Преподаватель:*

Кугаевских Александр Владимирович



Санкт-Петербург, 2023

**Модуль 2.**

**Лабораторная работа 1. Метод линейной регрессии**

**Введение**Изображение выглядит как текст, Шрифт, снимок экрана, документ

Автоматически созданное описание

**Описание метода**

Метод линейной регрессии - это статистический метод, используемый для определения связи между зависимой и независимыми переменными. Принцип работы метода заключается в построении линии наилучшего соответствия данных.

**Псевдокод метода**

1. Подготовить данные: разделить данные на тренировочный и тестовый наборы.

2. Выбрать модель: определить вид модели линейной регрессии.

3. Обучить модель: подобрать параметры модели с использованием тренировочных данных.

4. Оценить модель: оценить точность модели на тестовом наборе данных.

5. Применить модель: использовать обученную модель для прогноза значений зависимой переменной.

**Результаты выполнения**

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

**Примеры использования метода**

Метод линейной регрессии может быть полезен во многих ситуациях. Например, он может быть использован для прогнозирования продаж, основываясь на данных о рекламных затратах или других факторах, таких как время года, погода и т. д. Также, данный метод может быть применен для анализа взаимосвязей между различными переменными, таких как цена товара, его характеристики и спрос на рынке. Выбор метода линейной регрессии обусловлен его простотой и понятностью, а также хорошей прогностической способностью.

**Лабораторная работа 2. Метод k-ближайших соседей (k-NN)**

**Введение**

Изображение выглядит как текст, снимок экрана, Шрифт, документ

Автоматически созданное описание**Описание метода**

Метод k-ближайших соседей используется для классификации объектов на основе их близости к примерам обучающей выборки. Основной принцип работы метода заключается в нахождении k ближайших соседей объекта и отнесении его к классу, который наиболее часто встречается среди этих соседей.

**Псевдокод метода**

1. Для каждого объекта из обучающей выборки:
2. Вычислить расстояние между объектом из обучающей выборки и новым объектом.
3. Отсортировать объекты из обучающей выборки по возрастанию расстояния.
4. Выбрать k ближайших соседей.
5. Результирующее значение целевой переменной для нового объекта будет равно значению целевой переменной, которое имеют наиболее часто встречающиеся среди его k ближайших соседей.

**Результаты выполнения**

Изображение выглядит как диаграмма, линия, График, снимок экрана

Автоматически созданное описание Изображение выглядит как диаграмма, линия, График, дизайн

Автоматически созданное описание Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

**Примеры использования метода**

Метод k-ближайших соседей может быть полезен в следующих ситуациях:

Когда у нас есть обучающая выборка, для которой известны значения целевой переменной, и мы хотим классифицировать новый объект.

Когда данные имеют сложную структуру и требуют нелинейной модели для классификации или регрессии.

**Лабораторная работа 3. Деревья решений**

**Введение**

Изображение выглядит как текст, снимок экрана, Шрифт, документ

Автоматически созданное описание**Описание метода**

Метод деревьев решений является методом машинного обучения, который основывается на создании дерева, в котором каждый узел представляет условие или атрибут, а каждое ребро - результат этого условия. Дерево решений используется для прогнозирования или принятия решений на основе заданных данных. Он может применяться как для задач классификации, так и для задач регрессии.

**Псевдокод метода**

[A] Если все объекты в выборке относятся к одному классу, вернуть узел с этим классом

[B] Если все атрибуты уже рассмотрены, вернуть узел с наиболее часто встречающимся классом

[C] Иначе

[D] Найти атрибут с наибольшим приростом информации

[E] Создать узел для выбранного атрибута

[F] Для каждого значения атрибута создать потомок на дереве

[G] Рекурсивно применить алгоритм для новых потомков

**Результаты выполнения**

Изображение выглядит как текст, Шрифт, белый, чек

Автоматически созданное описаниеИзображение выглядит как линия, текст, График, диаграмма

Автоматически созданное описание

Изображение выглядит как диаграмма, линия, График, текст

Автоматически созданное описание**Примеры использования метода**

Метод деревьев решений может быть полезен в следующих ситуациях:

Классификация клиентов по их покупательскому поведению для оптимизации маркетинговых стратегий.

Прогнозирование вероятности оттока клиентов на основе их активности в приложении.

Определение причин аварий на основе данных о состоянии системы.

Решение задачи обнаружения мошеннических операций на основе истории транзакций.

**Лабораторная работа 4. Логистическая регрессия**

**Введение**

Изображение выглядит как текст, документ, снимок экрана, Шрифт

Автоматически созданное описание**Описание метода**

Логистическая регрессия - это метод машинного обучения, который используется для предсказания вероятности бинарного исхода на основе набора входных переменных. Он основан на логистической функции, которая преобразует линейную комбинацию входных переменных в вероятность отнесения к одному из классов.

**Псевдокод метода**

Инициализация:

- Инициализировать веса w случайными значениями

- Инициализировать смещение (bias) b нулевым значением

- Установить learning\_rate (скорость обучения)

- Установить количество итераций для обучения

Метод sigmoid:

- Принимает на вход массив x

- Возвращает 1 / (1 + exp(-x))

Метод predict:

- Принимает на вход данные x

- Вычисляет значения y\_pred путем применения функции sigmoid к (x \* w + b)

- Возвращает y\_pred

Метод train:

- Принимает на вход данные X и соответствующие метки y

- Для каждой итерации обучения:

- Вычисляет значения y\_pred с помощью метода predict для текущих весов w и смещения b

- Вычисляет градиенты для каждого веса и смещения с помощью формул градиентного спуска

- Обновляет веса w и смещение b, умножив градиенты на learning\_rate

- Возвращает обученные веса w и смещение b

**Результаты выполнения**

Изображение выглядит как текст, Шрифт, снимок экрана, белый

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание

**Примеры использования метода**

Метод логистической регрессии может быть полезен в различных ситуациях, например:

Классификация электронных писем как спам или не спам

Классификация покупателей как потенциальных или не потенциальных клиентов

Определение вероятности возникновения заболевания на основе набора медицинских параметров

**Сравнение методов**

**Сравнительный анализ методов**

**Линейная регрессия:**

- Преимущества: Простота реализации и интерпретации, хорошая производительность на данных с линейной зависимостью, подходит для предсказания непрерывных значений.

- Ограничения: Линейная регрессия не справляется с нелинейными зависимостями, чувствительна к выбросам и шуму в данных.

**Логистическая регрессия:**

- Преимущества: Хорошо работает при бинарной классификации, расчет вероятности принадлежности к классу, относительно проста для интерпретации.

- Ограничения: Плохо работает с данными, имеющими сложную нелинейную структуру, требует линейной разделимости классов.

**Деревья решений:**

- Преимущества: Может работать с любыми типами данных, обработка пропущенных значений и выбросов, хорошо интерпретируемый результат, легко обработать категориальные переменные.

- Ограничения: Склонны к переобучению на сложных данных, неустойчивость к небольшим изменениям в данных.

**Метод k ближайших соседей:**

- Преимущества: Простота реализации, хорошо работает на данных с нелинейной зависимостью, способен обрабатывать выбросы и шум в данных.

- Ограничения: Требуется хранение всего обучающего набора данных, неэффективен при работе с большими объемами данных, требуется определение и настройка значения k.

**Примеры лучшего использования каждого метода**

- Линейная регрессия может быть эффективна при предсказании цен на недвижимость, где зависимость между факторами и ценой может быть линейной.

- Логистическая регрессия может быть полезна для прогнозирования вероятности оттока клиентов в банковской отрасли или предсказания вероятности заболевания на основе медицинских данных.

- Деревья решений могут быть эффективны при принятии решений о предоставлении кредита, где нужно учитывать множество факторов.

- Метод k ближайших соседей может использоваться для классификации текстовых документов или обработки изображений с нелинейной структурой.

**Заключение**

В зависимости от типа данных, сложности задачи и требований к интерпретируемости, каждый из этих методов имеет свои преимущества и ограничения. Важно выбирать метод, который наиболее подходит для конкретной задачи и обучать модель с использованием оптимальных гиперпараметров.

**Приложения**

Код реализованных методов:

Изображение выглядит как шаблон, пиксель

Автоматически созданное описание