

ОТЧЁТ ПО ПРОЕКТУ

Выполнил: Гаськов Валерий

«Реализация линейной и логистической регрессии с нуля»

Введение

В рамках данного экзаменационного проекта была выполнена практическая реализация двух базовых алгоритмов машинного обучения — линейной регрессии и логистической регрессии без использования готовых моделей из библиотек машинного обучения.

Цель проекта — углублённо изучить математические основы регрессионных моделей, понять механизм обучения моделей с помощью градиентного спуска, а также закрепить навыки анализа данных и визуализации результатов.

Проект реализован в среде Python (Jupyter Notebook) с использованием библиотек numpy, pandas, matplotlib.

Датасет:

Использован публичный датасет `master.csv`

Источник: [Kaggle — Suicide Rates Overview 1985 to 2016]

(<https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016/discussion/570568>)

1. Линейная регрессия

1.1 Теоретическая основа

Линейная регрессия используется для моделирования зависимости между входным признаком X и целевой переменной y в виде линейной функции:

$$\hat{y} = wX + b$$

где:

- w — вес (коэффициент наклона),
- b — смещение (bias).

В качестве функции ошибки используется Mean Squared Error (MSE):

$$L = \frac{1}{N} \sum (y - \hat{y})^2$$

1.2 Градиентный спуск

Для обучения модели реализован Batch Gradient Descent.

Градиенты функции потерь по параметрам вычисляются вручную:

$$\frac{\partial L}{\partial w} = -\frac{2}{N} \sum X(y - \hat{y})$$

$$\frac{\partial L}{\partial b} = -\frac{2}{N} \sum (y - \hat{y})$$

На каждой итерации параметры w и b обновляются с использованием скорости обучения (learning rate).

1.3 Описание графиков (линейная регрессия)

График 1 — Scatter plot данных и линия регрессии

На графике изображены:

- реальные точки исходных данных (точки),
- линия линейной регрессии, полученная после обучения модели.

График демонстрирует, насколько хорошо линейная модель аппроксимирует исходные данные. Видно, что линия регрессии проходит через основное скопление точек, что говорит о корректной настройке модели.

График 2 — График функции потерь (MSE) от числа итераций

Данный график показывает изменение значения ошибки в процессе обучения.

Наблюдается:

- резкое уменьшение ошибки на начальных итерациях,
- последующая стабилизация, что свидетельствует о сходимости градиентного спуска.

Это подтверждает правильность выбора learning rate и корректную реализацию алгоритма.

2. Логистическая регрессия

2.1 Теоретическая часть

Логистическая регрессия применяется для задач бинарной классификации.

Модель использует сигмоидную функцию:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Выход модели интерпретируется как вероятность принадлежности к классу 1.

2.2 Функция потерь

Для логистической регрессии используется Log Loss (Binary Cross-Entropy):

$$L = -\frac{1}{N} \sum [y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

Градиенты функции потерь также вычисляются вручную, без использования готовых реализаций.

2.3 Описание графиков (логистическая регрессия)

График 3 — Сигмоидная функция

На графике показана S-образная кривая сигмоиды.

Она иллюстрирует:

- переход вероятности от 0 к 1,
- область неопределённости около значения 0.5,
- насыщение функции при больших положительных и отрицательных значениях.

График 4 — Разделяющая граница (decision boundary)

График демонстрирует:

- точки двух классов,
- линию принятия решения, построенную логистической регрессией.

Модель корректно разделяет классы, что указывает на успешное обучение и правильную реализацию алгоритма.

График 5 — График функции потерь логистической регрессии

На графике видно снижение значения log loss по мере обучения модели.

Это подтверждает:

- устойчивость градиентного спуска,
- адекватный подбор гиперпараметров,
- корректную математическую реализацию.

3. Анализ и результаты

В ходе выполнения проекта:

- обе модели были реализованы полностью с нуля;
- использовались только базовые математические операции;

- достигнута сходимость функций потерь;
- визуализации подтвердили корректность обучения.

Проект демонстрирует понимание:

- принципов оптимизации,
- математической основы регрессии,
- интерпретации графиков и результатов.

Заключение

В результате выполнения экзаменационного проекта были успешно реализованы линейная и логистическая регрессии без использования готовых ML-библиотек.

Проект позволил:

- углубить знания в области машинного обучения,
- закрепить работу с градиентным спуском,
- научиться анализировать поведение моделей через графики.

Полученные результаты подтверждают корректность реализации и соответствие поставленным целям.