

Лабораторная работа №3. Линейная регрессия нескольких переменных.

В данной работе вам предстоит реализовать линейную регрессию от нескольких переменных методом градиентного спуска с нормализацией данных, а также методом нормальных уравнений. В работе многое совпадает с заданием работы №2. Если в первой работе вы запрограммировали требуемые функции сразу в матричной форме, то выполнение этой работы будет легче.

Для выполнения работы запустите редактор (Visual Studio Code, PyCharm или другие) и откройте файл `lab03.py`

В этой работе необходимо построить и опробовать регрессионную модель для предсказания стоимости дома в зависимости от его площади и количества этажей.

Файл `ex1data2.txt` содержит обучающий набор данных из трех столбцов:

- площадь дома;
- количество этажей дома;
- стоимость дома.

1. Нормализация

Параметры x_1 (площадь, первый столбец) и x_2 (количество этажей, второй столбец) очень сильно отличаются по своим значениям, поэтому перед обучением модели выполним нормализацию выборки.

В программе имеется заготовка функции `feature_normalize` (ищите после функции `main`), которая получает исходную матрицу X , а возвращает нормализованную матрицу X_{norm} , а также параметры, при которых осуществлялась нормализация: μ (среднее значение), σ (стандартное отклонение).

Напишите содержимое функции `feature_normalize` для выполнения нормализации. Теоретический материал по нормализации дан в разделе 2.2 учебника и лекции №3.

2. Градиентный спуск

На данном шаге вам необходимо написать функцию стоимости и градиентного спуска для нескольких переменных. В программе есть заготовки `compute_cost` и `gradient_descent`. Если в первой части вы правильно реализовали эти функции сразу в матричной форме, то на данном шаге ничего менять не придется, можно скопировать решение из прошлой части.

Когда функции готовы, запустите программу. Она вызовет алгоритм градиентного спуска для обучения модели при различных значениях параметра скорости спуска α , а затем отобразит графики зависимости функции стоимости от шага алгоритма для каждого из вариантов α .

Оцените графики и найдите значение α , наиболее удачное для обучения модели.

Наконец, полученная модель используется для предсказания стоимости дома с площадью и количеством этажей, которых нет в обучающей выборке.

Дополнительно: Придумайте несколько своих вариантов и получите для них предсказание стоимости, добавив в программу еще примеры.

3. Нормальные уравнения

На последнем шаге вам необходимо решить ту же задачу линейной регрессии, но уже методом нормальных уравнений. Теоретический материал по нормальным

уравнениям обсуждается в разделе 2.4 учебника и лекции №3.

В программе имеется заготовка функции `normal_equation`, которая принимает на вход обучающую выборку и возвращает оптимальные значения *theta* для линейной регрессии.

Напишите содержимое функции `normal_equation` и запустите программу. Обратите внимание, что для обучения модели исходные данные берутся без нормализации.

Полученные из функции `normal_equation` значения *theta* используются для предсказания стоимости дома с теми же параметрами, что и в варианте для градиентного спуска. Сравните эти значения (а также придуманные вами варианты, если были). Предсказанные стоимости домов методами градиентного спуска и нормальных уравнений должны быть близки.

На этом выполнение лабораторной работы №3 завершается.