МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Институт №8 «Информационные технологии и прикладная математика»

Кафедра 810 «Информационные технологии в моделировании и управлении»

**Лабораторная работа №1**

**по курсу «Интеллектуальный анализ данных»**

**Метод градиентного спуска. Стохастический градиентный спуск. Метод Momentum..**

Выполнил: А.С.Бобряков

Группа: М8O-103М-19

Преподаватель: К.К.Абгарян

Москва, 2020

**Задание**

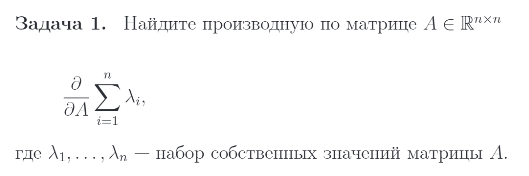
Реализовать линейную регрессию с функцией потерь MSE, обучаемую с помощью:

1. метода градиентного спуска (упрощенный вариант);

2. метода стохастического градиентного спуска;

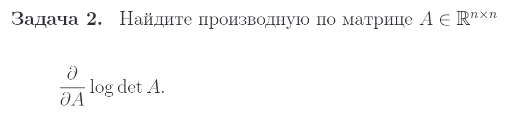
3. метода Momentum;

**Ход выполнения работы**

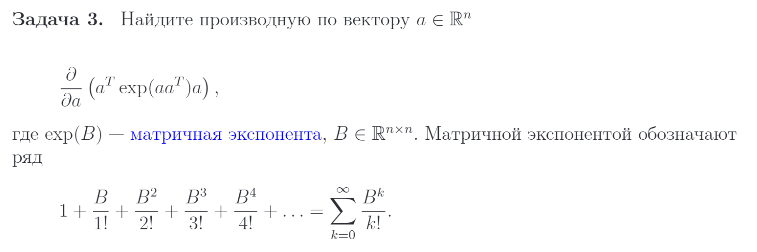


**Решение:**

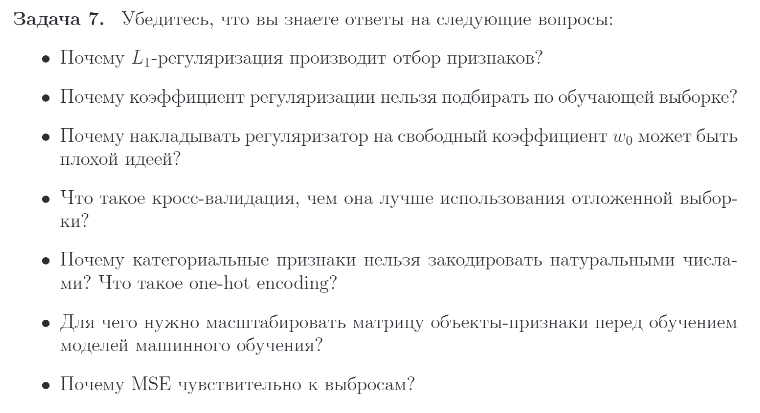
где Tr(A) – след матрицы A. Источники: Свойства следа матрицы <https://ru.wikipedia.org/wiki/След_матрицы> и свойства дифференцирования следа матицы по матрице <http://mas.exponenta.ru/literature/Spur1.pdf>.



**Решение:**



**Решение:**



1) Потому что L1 норма зануляет некоторые веса, если предсказательная способность вектора мала. То есть происходит отсечение неинформативных признаков.

2) Потому что регуляризация борется с переобучением, когда модель показывает значительно меньшую предсказательную способность на обучающей выборке чем на тестовой.

3) Если мы будем штрафовать за его величину, то получится, что мы учитываем некие априорные представления о близости целевой переменной к нулю и отсутствии необходимости в учёте её смещения.

4) Кросс-валидация - процедура эмпирического оценивания обобщающей способности алгоритмов, обучаемых по прецедентам. Фиксируется некоторое множество разбиений исходной выборки на две подвыборки: обучающую и контрольную. Кросс-валидация лучше отложенной выборки, т.к. в данной процедуре циклы кросс-валидации проводятся на разных разбиениях, а результаты валидации усредняются по всем циклам, что повышает качество проверки.

5) Основной проблемой кодирования категориальных признаков натуральными числами является то что числовой код создает евклидово представление для данных и неявным образом вводится алгебра над значениями этих признаков, которая не имеет смысла. При такой кодировке не применимым линейные модели.

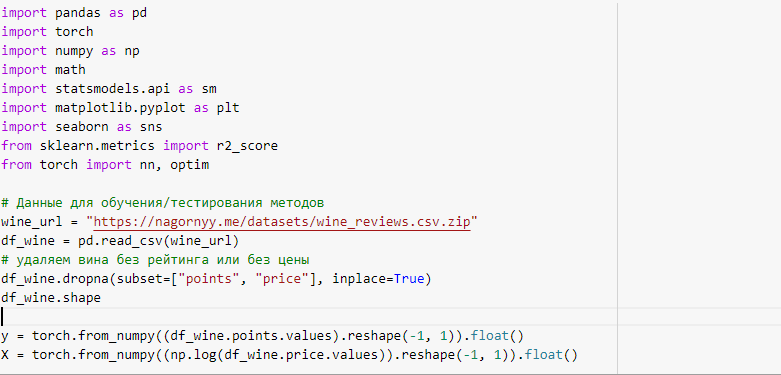
One Hot Encoding подразумевает создание n признаков, все из которых равны нулю за исключением одного. На позицию, соответствующую численному значению признака мы помещаем 1.

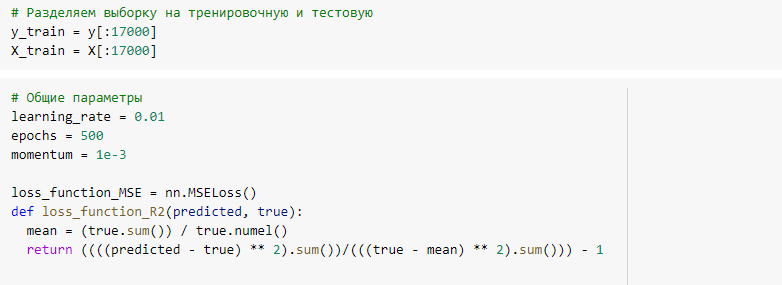
6) Масштабирование признаков означает приведение их к одному диапазону значений, т.к. многие методы некорректно работаю для различных диапазонов разных свойств в одном наборе данных.

7) MSE чувствителен к выбросам потому что является усредненным значением, и будет сильно смещена если хотя бы одно из значений будет значительно больше или меньше остальных. Такое может произойти, в частности, если не было проведено масштабирование признаков.

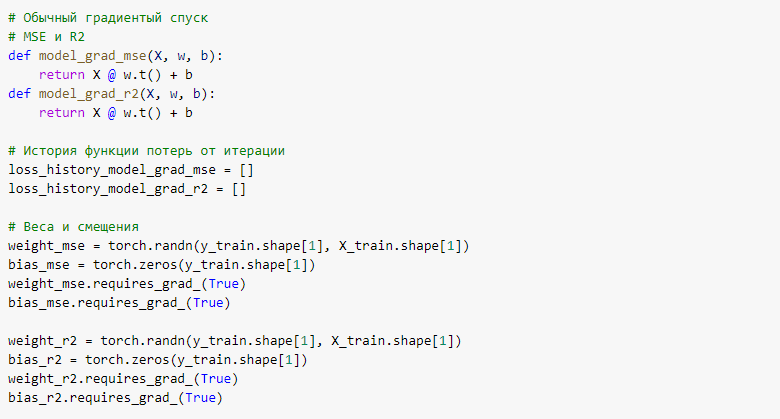
**Ход решения лабораторной работы.**

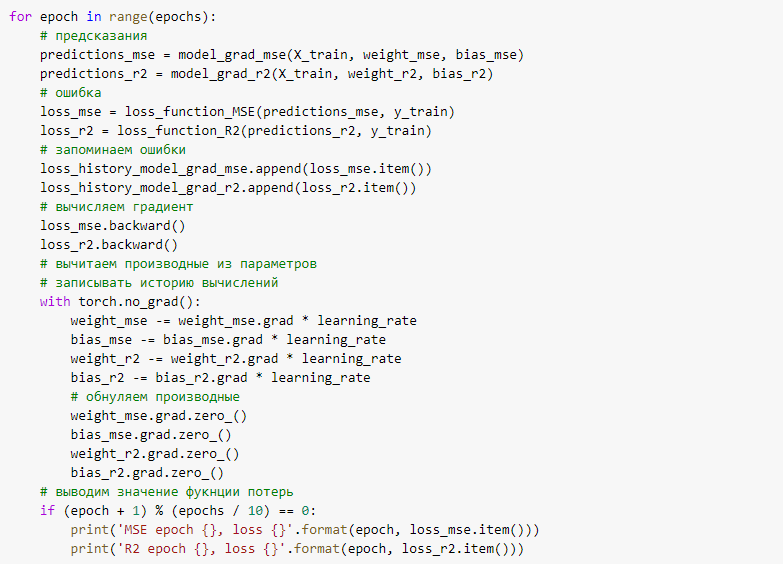
Загрузка данных (Задание 4)

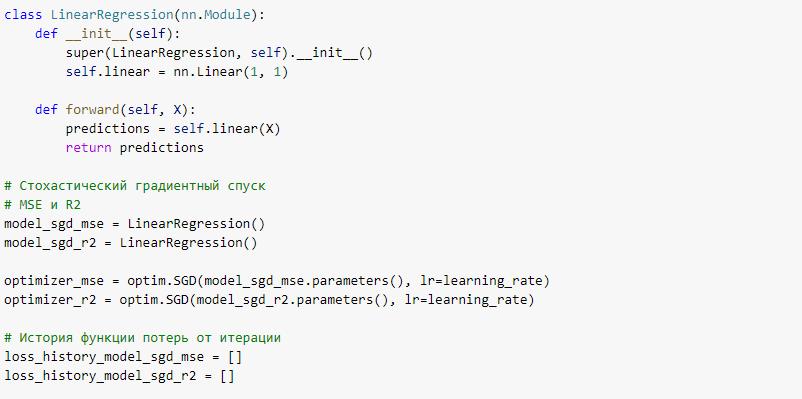


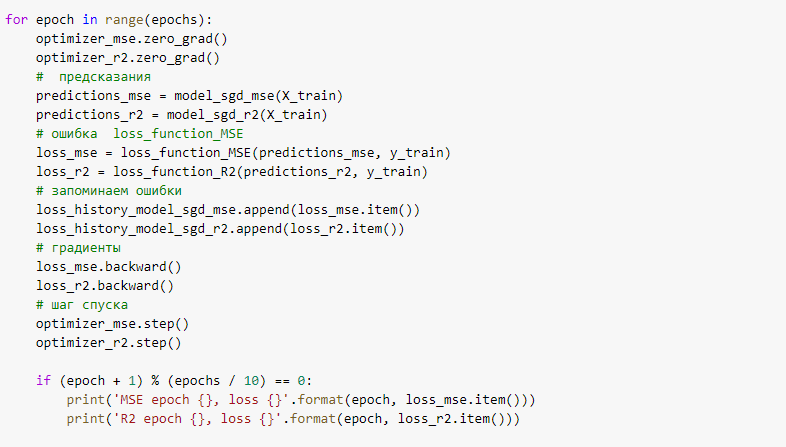


Реализация Градиентного спуска + Стохастического градиентного спуска + метода Momentum + Реализация в них MSE + R2 (Задания 1, 2, 3, 5)



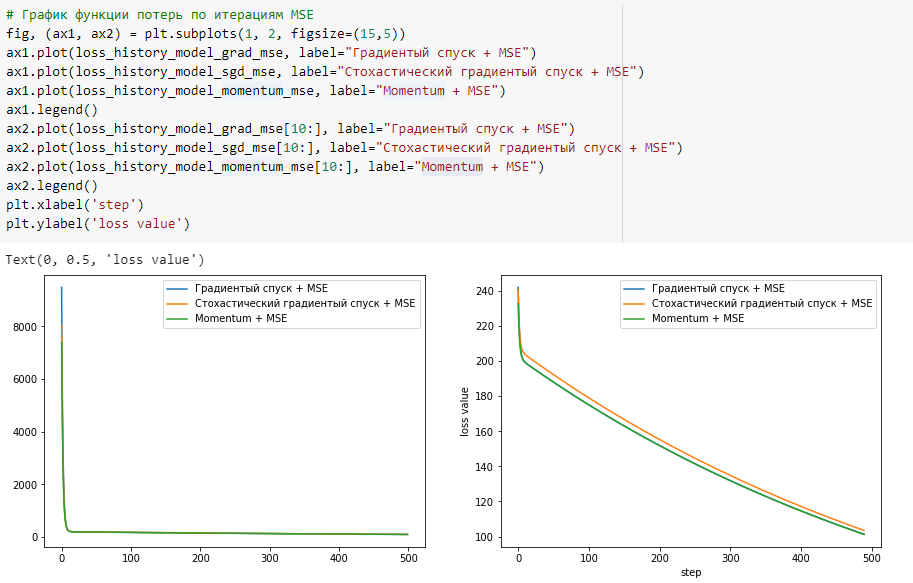


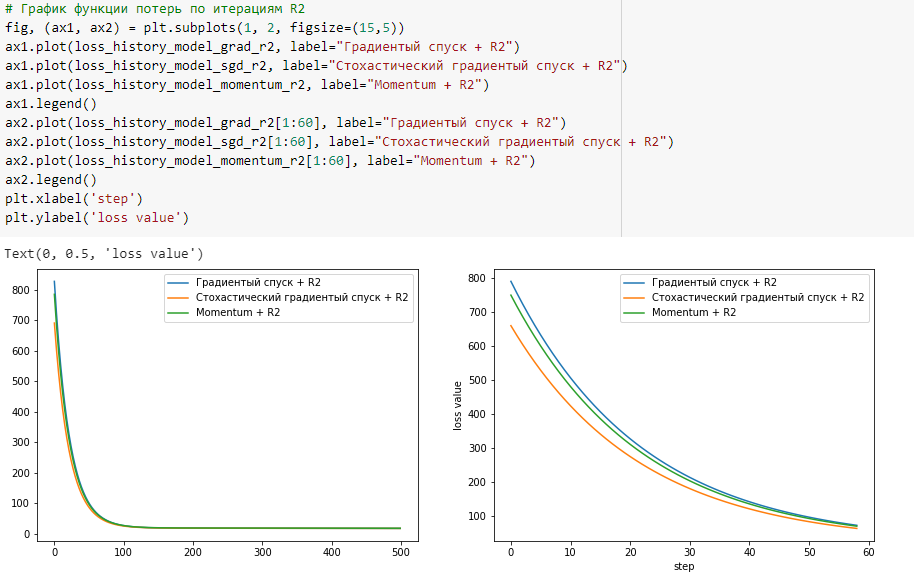






По итогам обучения построим функции зависимости значения MSE/R2 от шага итерации:





**Выводы**

В ходе лабораторной работы были изучены и запрограммированы методы градиентного спуска и его модификации. Построены графики обучения моделей для различных методов с использованием функций потерь MSE и R2.