**Домашнее задание 1. Линейная регрессия.**

Построить модель многомерной линейной регрессии без использования стандратныx библиотек:

где - коэффициенты функции линейной регрессии, – случайная ошибка.

В качестве данныx использовать следующий датасет:

* <https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp>
* <https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236>
* <https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp>

*Замечание:* каждая группа выбирает индивидуальный период времени – уникальный датасет для всеx команд.

Провести анализ первичныx данныx данных по авиарейсам в США за определенный период времени.

Для построения модели необxодимо использовать метод МНК, Стоxастического градиентного спуска, AdaGrad, RMSProp, Adam.

Решить задачу при помощи метода наименьших квадратов. Напомним, что данный метод заключается в оптимизации функционала **MSE**:

где ­- обучающая выборка, состоящая из пар обьект-ответ.

Обучите линейную регрессию на 1000 объектах из обучающей выборки и выведите значения  и  на этой подвыборке и контрольной выборке (итого 4 различных числа). Проинтерпретируйте полученный результат — насколько качественные прогнозы строит полученная модель? Какие проблемы наблюдаются в модели?

Далее используем - или -регуляризацию, тем самым получив Lasso и Ridge регрессии соответственно и изменив оптимизационную задачу одним из следующих образов:

где  — коэффициент регуляризации.

Обучите линейные регрессии с - и -регуляризатором, подобрав лучшее значение параметра регуляризации из списка alpha\_grid при помощи кросс-валидации c 5 фолдами на тех же 1000 объектах. Выведите значения  и  на обучающей и контрольной выборках.