

# Линейная регрессия

## Цели работы:

- 1) реализовать два способа решения задачи линейной регрессии;
- 2) настроить гиперпараметры у каждого алгоритма, в частности параметры одного из методов регуляризации;
- 3) анализ результатов.

## Данные

Используйте один из [этих наборов данных](#) для тестирования алгоритмов. Каждый тест в архиве организован следующим образом:

```
%число признаков%
%число объектов в тренировочном наборе%
%объект тренировочного набора 1%
%объект тренировочного набора 2%
.....
%объект тренировочного набора N%
%число объектов в тестовом наборе%
%объект тестового набора 1%
%объект тестового набора 2%
.....
%объект тестового набора K%
```

Формат объектов совпадает с форматом из соответствующей задачи на Codeforces.

## Задание

### Алгоритмы

Реализуйте алгоритмы нахождения уравнения прямой для задачи линейной регрессии:

- МНК — метод наименьших квадратов (псевдообратная матрица / SVD);
- градиентный спуск.

*На лекции мы рассматривали алгоритм градиентного спуска для классификации, однако его можно применять и для задач регрессии, важно лишь выбрать дифференцируемую функцию ошибки. В данном случае необходимо использовать среднюю квадратичную ошибку.*

Требуется реализовать стохастический или пакетный градиентный спуск. Напоминаем, что эмпирический риск нужно балансировать на каждой итерации при помощи экспоненциального скользящего среднего.

Для алгоритма градиентного спуска рекомендуется использовать начальную инициализацию весов  $w_i \in [-\frac{1}{2n}; \frac{1}{2n}]$ , где  $n$  — число признаков (см. лекцию). Шаг градиента необходимо уменьшать на каждой итерации, например:  $\mu_k = \frac{1}{k}$ ,  $k$  — номер итерации. Другие способы инициализации весов и уменьшения шага градиента использовать также *можно*.

Алгоритм градиентного спуска необходимо запустить с ограничением по числу итераций (не более 2000 итераций).

В качестве функции оценки качества алгоритма используйте **NRMSE**, либо **SMAPE**.

## Регуляризация

В реализации каждого из вышеупомянутых алгоритмов необходимо использовать регуляризацию. Для МНК гребневую регуляризацию, для градиентного один из методов на выбор:

- гребневая;
- LASSO;
- Elastic Net.

## Настройка и анализ

Для каждого алгоритма найдите наилучшие гиперпараметры, а именно, параметры регуляризации, и выведите лучшие соответствующие результаты с точки зрения выбранной Вами функции ошибки. Перебирать различные способы инициализации вектора весов, уменьшения шага градиента а также различные темпы затухания в экспоненциальном скользящем среднем в качестве гиперпараметров **не требуется**.

Для алгоритма градиентного спуска постройте график зависимости функции оценки качества (**NRMSE** или **SMAPE**) на тренировочном и тестовом множестве от числа итераций.