

# Домашнее задание по курсу «Методы машинного обучения»: Отбор и генерация признаков

Терлиженко Ирина

15 февраля 2021 г.

Необходимо реализовать и исследовать на практике несколько feature selection алгоритмов. В качестве датасета использовать Boston dataset . В качестве предсказываемого признака (метки) используется признак MEDV. При необходимости используйте нормализацию признаков.

В задании 1 все вычисления (кроме скачивания датасета) необходимо делать без использования каких-либо ML библиотек (sklearn, pandas можно использовать только для визуализации). В остальных задачах разрешено использование библиотек кроме методов, реализующих непосредственно feature selection.

## 1 Задания

1. (**+2 балла**) Выберите 5 лучших признаков на основании Filter метода. В качестве меры качества используйте коэффициент корреляции Пирсона.
2. Выберите 5 лучших признаков на основании Forward (**+1 балл**) и Backward (**+1 балл**) Greedy Selection методов. В качестве базового алгоритма используйте линейную регрессию с квадратичной функцией потерь.
3. (**+2 балл**) Выберите 5 лучших признаков, используя линейную регрессию с квадратичной функцией потерь и регуляризацией Лассо. В отчете опишите, как вы выбирали параметр регуляризации  $\lambda$ .
4. (**Bonus: +1 балл**) Докажите, что

$$\min_{a,b \in R} \frac{1}{m} \|av + b - y\|^2 = \min_{a,b \in R} \frac{1}{m} \|a(v - \bar{v}) + b - (y - \bar{y})\|^2$$

Совет: пусть  $a^*, b^*$  — значения, минимизирующие левую часть. Подберите такие  $a, b$ , что правая часть не больше левой. Аналогично в обратную сторону.