

Регуляризация линейной регрессии

Лассо, или L1-регуляризация

Skillbox

образовательная платформа

L1-регуляризация — она же «манхэттенское расстояние», или «регрессия лассо».

LASSO (Least absolute shrinkage and selection operator — оператор наименьшего абсолютного сокращения и выбора).

Основная идея заключается в том, чтобы оставить в модели только наиболее важные переменные, которые влияют на конечный результат.

То есть L1-регуляризация выглядит как способ отбора признаков.

Если основная формула ошибки линейной регрессии, когда мы должны минимизировать квадратичную функцию потерь, следующая:

$$L = \sum_{i=1} (y_i - f(x_i))^2 \rightarrow \min$$

То в случае L1-регуляризации формула принимает вид

$$L = \sum_{i=1} (y_i - f(x_i))^2 + \lambda \sum_{i=1} |w_i| \rightarrow \min,$$

где w_0, w_1, \dots, w_k — параметры модели

