Линейная регрессия

Лекция 2

План лекции

- Понятие линейных моделей
- Измерение ошибки в задачах регрессии
- Обучение линейной регрессии
- Градиентный спуск
- Стохастический градиентный спуск
- Модификации градиентного спуска
- Предобработка данных
- Переобучение
- Оценка качества моделей
- Регуляризация

Понятие линейных моделей

- Линейная регрессионная модель:

$$a(\vec{x}_i) = w_0 + \sum_{j=1}^{d} w_j x_{ij}\ ,$$

где w_j – веса или весовые коэффициенты,
w_0 – свободный коэффициент или сдвиг (bias).

- В векторном виде:

$$a(\vec{x}_i) = w_0 + \langle \vec{w}, \vec{x}_i \rangle,$$

где $\vec{w} = (w_1, \dots, w_d)$, $\vec{x}_i = (x_{i1}, \dots, x_{id})$.

- В сокращенном векторном виде:

$$a(\vec{x}_i) = \langle \vec{w}, \vec{x}_i \rangle$$

Измерение ошибки в задачах регрессии

- Функция потерь:

$$L(y, y_{pred}) = L(y, a)$$

- Среднеквадратичная ошибка (mean squared error, MSE):

$$L(y, a) = (a - y)^2$$

$$MSE(a, X) = \frac{1}{l}\sum_{i=1}^{l} L(y_i, a_i) = \frac{1}{l}\sum_{i=1}^{l} (a(\vec{x}_i) - y_i)^2$$

- Root mean squared error (RMSE):

$$RMSE(a, X) = \sqrt{\frac{1}{l}\sum_{i=1}^{l} (a(\vec{x}_i) - y_i)^2}$$

Измерение ошибки в задачах регрессии

- Коэффициент детерминации:

$$R^2(a, X) = 1 - \frac{\sum_{i=1}^{l}(a(\vec{x}_i) - y_i)^2}{\sum_{i=1}^{l}(y_i - \bar{y})^2} = 1 - \frac{\sigma^2}{\sigma_y^2}$$

где σ_y^2 – дисперсия y, σ^2 – дисперсия ошибки модели

- Среднее абсолютное отклонение (mean absolute error, MAE):

$$L(y, a) = |a - y|$$

$$MAE(a, X) = \frac{1}{l}\sum_{i=1}^{l}|a(\vec{x}_i) - y_i|$$

Измерение ошибки в задачах регрессии

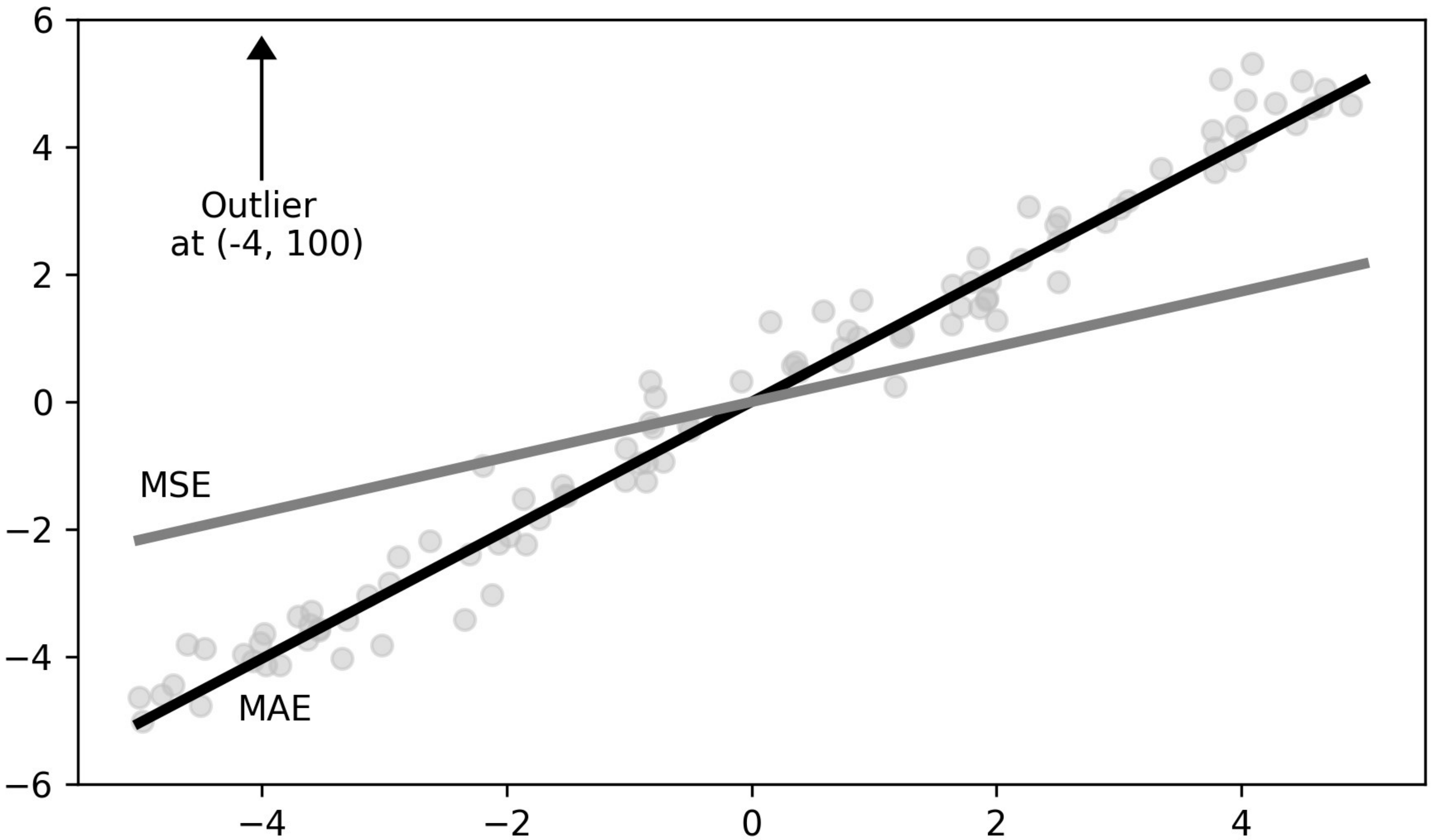
- Среднеквадратичная логарифмическая ошибка (mean squared logarithmic error, MSLE):

$$L(y, a) = \big(log(a + 1) - log(y + 1)\big)^2$$

- Средняя абсолютная процентная ошибка (mean absolute percentage error, MAPE):

$$L(y, a) = \left|\frac{y - a}{y}\right|$$

Измерение ошибки в задачах регрессии



Обучение линейной регрессии

- В случае использования среднеквадратичной ошибки (MSE):

$$\frac{1}{l}\sum_{i=1}^{l}(\langle\vec{w},\vec{x}_i\rangle - y_i)^2 \to \min_{\vec{w}}$$

- В матричном виде:

$$\frac{1}{l}\|X\vec{w} - \vec{y}\|^2 \to \min_{\vec{w}},$$

где $X \in \mathbb{R}^{l\times d}$, $\vec{w} \in \mathbb{R}^d$, $\vec{y} \in \mathbb{R}^l$

Обучение линейной регрессии

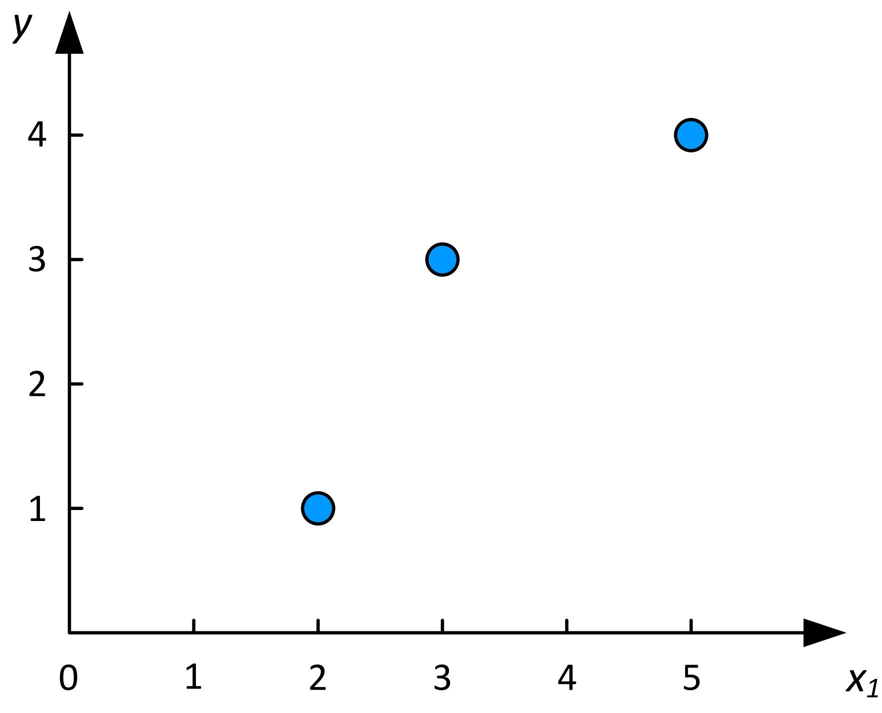
- После дифференцирования данного функционала по вектору $\vec{w}$, приравнивания к нулю и решения уравнения, получаем:

$$\frac{\partial}{\partial \vec{w}}\left(\frac{1}{l}\|X\vec{w}-\vec{y}\|^2\right)=0 \ \rightarrow \vec{w}_{opt}=(X^T X)^{-1}X^T\vec{y}$$

– *нормальное уравнение* (normal equation)

Обучение линейной регрессии

- Пример: пусть даны три точки $(2, 1)$, $(3, 3)$, $(5, 4)$
- Требуется построить линейную регрессионную модель на основе нормального уравнения



Обучение линейной регрессии

$$X = \begin{array}{c} \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix} \\ \begin{array}{cc} x_0 & x_1 \end{array} \end{array}, \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

Обучение линейной регрессии

$$X = \begin{array}{c} \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix} \\ \begin{matrix} x_0 & x_1 \end{matrix} \end{array}, \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix} =$$

Обучение линейной регрессии

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}, \qquad \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

$$x_0 \quad x_1$$

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 10 \\ 10 & 38 \end{bmatrix}$$

$$2 \times 3 \qquad 3 \times 2 \qquad 2 \times 2$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} =$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 2.7 & -0.7 \\ -0.7 & 0.2 \end{bmatrix}$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 2.7 & -0.7 \\ -0.7 & 0.2 \end{bmatrix}$$

$$(X^T X)^{-1} X^T =$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 2.7 & -0.7 \\ -0.7 & 0.2 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \begin{bmatrix} 1.29 & 0.57 & -0.85 \\ -0.28 & -0.07 & 0.36 \end{bmatrix}$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 2.7 & -0.7 \\ -0.7 & 0.2 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \begin{bmatrix} 1.29 & 0.57 & -0.85 \\ -0.28 & -0.07 & 0.36 \end{bmatrix}$$

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y} =$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 2.7 & -0.7 \\ -0.7 & 0.2 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \begin{bmatrix} 1.29 & 0.57 & -0.85 \\ -0.28 & -0.07 & 0.36 \end{bmatrix}$$

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y} = \begin{bmatrix} -0.43 \\ 0.93 \end{bmatrix} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

Обучение линейной регрессии

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y}$$

$$(X^T X)^{-1} = \begin{bmatrix} 2.7 & -0.7 \\ -0.7 & 0.2 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \begin{bmatrix} 1.29 & 0.57 & -0.85 \\ -0.28 & -0.07 & 0.36 \end{bmatrix}$$

$$\vec{w}_{opt} = (X^T X)^{-1} X^T \vec{y} = \begin{bmatrix} -0.43 \\ 0.93 \end{bmatrix} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$a(\vec{x}) = w_0 + w_1 x_1 = -0.43 + 0.93 x_1$$

Обучение линейной регрессии

$$a(\vec{x}) = -0.43 + 0.93x_1, \qquad a(1) = 0.5, \qquad a(5) = 4.2$$