Задача №1

$y(x) = \theta x$, $x \in \mathbb{R}$, $\theta \in \mathbb{R}$ $Y_i = \theta x_i + \varepsilon_i$, $i \in \{1, \dots, n\}$

x_i - неслучайные

ε_i - случайные

1) $L_1 = \sum_{i=1}^{n} (Y_i - \theta x_i)^2$

$L_1 \to \min\limits_{\theta \in \mathbb{R}}$

Для нахождения $\hat{\theta}$ найдем $\min L$

$$\frac{dL_1}{d\theta} = \frac{d}{d\theta} \sum_{i=1}^{n} (Y_i - \theta x_i)^2 = \sum_{i=1}^{n} -2x_i (Y_i - \theta x_i) = -2\left(\sum_{i=1}^{n} x_i Y_i - \theta \sum_{i=1}^{n} x_i^2\right)$$

$$\frac{dL_1}{d\theta} = 0 \Rightarrow \sum_{i=1}^{n} x_i Y_i - \hat{\theta} \sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$

2) GD и SGD

- GD на каждой итерации i θ изменяется по формуле:

$$\theta_i = \theta_{i-1} - \alpha \cdot \nabla L(\theta_{i-1})$$

ф-ция потерь

learning rate

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \theta x_i)^2$$ - среднеквадратичная ошибка

$$\nabla L(\theta) = -\frac{2}{n} \sum_{i=1}^{n} x_i (Y_i - \theta x_i)$$ - градиент

Формула итераций:

$$\theta_i = \theta_{i-1} + \alpha \cdot \frac{2}{n} \sum_{k=1}^{n} x_k (Y_k - \theta_{i-1} x_k)$$, где n - размер всей выборки

◦ SGD

Для SGD формула такая:

$$\theta_i = \theta_{i-1} + \alpha \cdot \frac{2}{B}\sum_{k=1}^{B} x_k(Y_k - \theta_{i-1} x_k),$$ где B - размер батча

Батч - выборка случайных объектов из общего набора.

При этом батчи можно объявить до обучения, разбив всю выборку на $\frac{n}{B}$ батчей размером B.

Задача 3

$X \in \mathbb{R}^{n\times d}$, $Y \in \mathbb{R}^{n}$

$\theta \in \mathbb{R}^d$

$$\|Y - X\theta\|^2 + \lambda\|\theta\|^2 \to \min_\theta$$

1) θ - ?

$$L(\theta) = |Y - X\theta|^2 + \lambda|\theta|^2$$

$$\nabla L = -2X^T(Y - X\theta) + 2\lambda\theta$$

$$\nabla L = 0 \Rightarrow \theta = (X^TX + \lambda I)^{-1}X^TY$$

единичная матрица

В МНК $\theta = (X^TX)^{-1}X^TY$

Разница только в слагаемом λI, которое решает проблему обратимости матрицы X^TX

2) GD

$$\nabla L(\theta) = -2X^T(Y - X\theta) + 2\lambda\theta$$

$$\theta_i = \theta_{i-1} - \alpha\nabla L(\theta_{i-1})$$

SGD

$$\nabla L_i(\theta) = -2 x_i^T (y_i - x_i\theta) + 2\lambda\theta$$

$$\theta_k = \theta_{k-1} - \alpha \nabla L_i(\theta_{k-1})$$

Батч определяется как независимый набор случайных объектов из общего набора, где случайные величины равномерно распределены

3) Стандартизация важна, так как регуляризационный член штрафует модель за большие коэффициенты. Если признаки имеют разные масштабы, то это может привести к тому, что некоторые коэффициенты будут штрафоваться сильнее остальных, соответственно, их вес будет искажен.