$$y = X \cdot \beta + \varepsilon \quad (n,1)$$

$$(n,1) \quad (n, k+1) \cdot (k+1, 1)$$

$$(n, 1)$$

L$_2$-норма вектора (Чебышева)

$$v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \|v\|_2^2 = v^T \cdot v = (1 \;\; 2) \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1^2 + 2^2$$

$$v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad w = \begin{pmatrix} 3 \\ 4 \end{pmatrix} \qquad (v^T \cdot w)^T = w^T \cdot v = v^T \cdot w = 1 \cdot 3 + 2 \cdot 4$$

$$(X\beta)^T = \beta^T X^T$$

$$\|\varepsilon\|_2^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

$$\|y - X\beta\|^2 = \sum_{i=1}^{n} (y_i - x_i \beta)^2$$

$$error = (y - X\beta)^T (y - X\beta) = y^T y - \underbrace{\overset{w^T}{y^T} \overset{v}{(X\beta)}} - \underbrace{\overset{v^T}{(X\beta)^T} \overset{w}{y}} + (X\beta)^T \cdot X\beta =$$

$$= y^T y - 2\beta^T \cdot X^T y + \beta^T X^T X \beta$$

$$(k+1, n) \cdot (n, 1) \qquad (k+1, n) \cdot (n, k+1)$$

$$\to (k+1, 1) \qquad \to (k+1, k+1)$$

$$\nabla_\beta error = \frac{\partial\, error}{\partial \beta} = -2X^T y + 2X^T X \beta = 0$$

$$(k+1, n)\,(n, 1) \qquad (k+1, k+1) \cdot (k+1, 1)$$

$\to (k+1, 1)$ $\qquad\qquad$ $\to (k+1, 1)$

размерность производной совпадает с аргументом

Производные скалярных функций

$$(x\cdot 5)'_x = 5 \qquad (x\cdot 5\cdot x)'_x = 2\cdot 5\cdot x$$

матричные производные:

$$(x^T\cdot V)'_x = V \qquad (x^T A x)'_x = 2Ax$$

$$-2X^Ty + 2X^TX\beta = 0$$

$$\cancel{2}X^Ty = \cancel{2}X^TX\beta$$

$$X^Ty = X^TX\beta$$

$$(X^TX)^{-1}X^Ty = \underbrace{(X^TX)^{-1}X^TX}_{=I}\beta \Rightarrow \hat{\beta} = (X^TX)^{-1}\cdot X^Ty$$

ДЗ:

$$\text{error} = (y - X\beta)^T(y - X\beta) + \lambda\cdot\beta^T\cdot I\cdot\beta$$

$$\varepsilon \sim N(0, \sigma^2 \cdot I)$$

$$y = X\beta + \varepsilon$$

$$y \sim N(X\beta, \sigma^2 \cdot I).$$

$$x \sim N(\mu, \Sigma)$$

$$p(x) = (2\pi)^{-\frac{n}{2}} \cdot (\det(\Sigma))^{-1} \cdot \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right)$$

$$\downarrow$$

$$p(y) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\sigma^2 I)}} \cdot \exp\left(-\frac{1}{2\det(\sigma^2 I)} \|y - X\beta\|^2\right)$$

$$\log(p(y)) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\underbrace{\|y-X\beta\|^2}$$

$$MSE = \frac{1}{n}\|y-X\beta\|^2$$

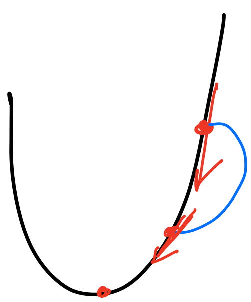
$$\log \mathcal{L}(\beta) \to \max_\beta \iff MSE(\beta) \to \min_\beta$$

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = 1\cdot 1 - 0\cdot 0 = 1$$

$$\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = 1\cdot\begin{vmatrix}1 & 0\\ 0 & 1\end{vmatrix} = 1$$

$$\left|\begin{pmatrix}3 & 0\\ 0 & 3\end{pmatrix}\right| = 3\cdot 3 - 0\cdot 0 = 3^2$$

$$\nabla_\beta error = \frac{\partial\, error}{\partial \beta} = -2X^T y + 2X^T X\beta$$



MSE

$-\nabla_\beta error$

$\beta_0$

шаг обучения

$\beta_1 = \beta_0 + \alpha\cdot\nabla_\beta error$

$\beta_2 = \beta_1 + \alpha\cdot\nabla_\beta error$