$$d f(x) = \underbrace{f'_x(x) \cdot dx} \qquad \underset{2\times 1}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

$$\to \; df(x_1, x_2) = \underbrace{f'_{x_1}(x)\, dx_1 + f'_{x_2}(x)\, dx_2}$$

$$df = \underbrace{\begin{bmatrix} dx_1 & dx_2 \end{bmatrix}}_{=:dx^T} \underbrace{\begin{bmatrix} f'_{x_1} \\ f'_{x_2} \end{bmatrix}}_{grad} = dx^T \cdot grad(x_1, x_2)$$

$$df = dx^T \cdot grad(x_1, x_2)$$

1. $dx^T \cdot \underbrace{\boxed{}}_{\text{градиент того, что диф-м}}$

$$A = \begin{pmatrix} x & x^2 \\ 2x & 3 \end{pmatrix} \qquad dA = \begin{pmatrix} dx & 2x \cdot dx \\ 2\,dx & 0 \end{pmatrix}$$

2. $d(AB) = \underbrace{dA \cdot B} + \underbrace{A \cdot dB}$

R, r – мат., век., зав. от перем.

A, B – мат. констант.

1. $d(A \cdot R \cdot B) = A \cdot dR \cdot B$

2. $d(r^T \cdot r) = \underbrace{\underbrace{dr^T}_{1\times n} \cdot \underbrace{r}_{n\times 1}}_{1\times 1} + \underbrace{\underset{1\times n}{r^T} \cdot \underset{n\times 1}{dr}}_{1\times 1} =$

$\underset{n\times 1}{r}$

$\underset{1\times 1}{a} = \underset{1\times 1}{a^T}$

$= dr^T \cdot r + dr^T \cdot r = \underset{1\times n}{dr^T} \cdot \underset{n\times 1}{2r}$

$\text{grad}(r^T r) = 2 \cdot r$

$\downarrow$

$(r^2)'_r = 2 \cdot r$

3. $d(r^T A r) = \underset{1\times n}{dr^T} \cdot \underset{n\times n}{A} \cdot \underset{n\times 1}{r} + \underset{1\times n}{r^T} \cdot \underset{n\times n}{A} \cdot \underset{n\times 1}{dr} =$

$\underset{n\times 1}{r} \quad \underset{n\times n}{A}$

$= dr^T A r + dr^T A^T r =$

$\to \quad = dr^T \underbrace{(A + A^T) r}$

$(a \cdot r^2)'_r = 2ar$

$= \text{grad}(r^T A r)$

4. $d\left(\cos(r^T r)\right) = dr^T \left(2r \cdot -\sin(r^T r)\right)$

$$MSE = \|y - \hat{y}\|_2^2 = \sum_i (y_i - \hat{y}_i)^2$$

$$\left[MSE = (y - X\hat{w})^T (y - X\hat{w})\right]'_{\hat{w}}$$

$\text{grad}(MSE)_{\hat{w}}$

$$dMSE = d\left[y^T y - \hat{w}^T X^T y - y^T X \hat{w} + \hat{w}^T X^T X \hat{w}\right] =$$

$$= -d\hat{w}^T X^T y - y^T X\, d\hat{w} + d\hat{w}^T X^T X \hat{w} +$$

$$+ \hat{w}^T X^T X\, d\hat{w} = -2\, d\hat{w}^T X^T y + 2\, d\hat{w}^T X^T X \hat{w} =$$

$$= d\hat{w}^T \underbrace{\left(-2X^T y + 2X^T X \hat{w}\right)}_{\text{grad}(MSE)_{\hat{w}}}$$

$$-\cancel{2}X^T y + \cancel{2}X^T X \hat{w} = 0$$

$$X^T X \hat{w} = X^T y$$

$$\hat{w} = (X^T X)^{-1} X^T y$$