$y_{m\times 1}$, $x_{n\times 1}$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

---

1. $\underset{m\times 1}{y} = \underset{m\times n}{A}\ \underset{n\times 1}{x}$

$$\frac{\partial y}{\partial x} = A$$

$$y_i = \sum_{i=1}^{n} a_{ik} x_k$$

$$\frac{\partial y_i}{\partial x_j} = a_{ij}$$

$$\frac{\partial y}{\partial x} = A$$

$$\underline{\overset{1\times m}{y}{}^{T}\overset{m\times n}{A}\overset{n\times 1}{x}} = \mathcal{L}$$

$$\frac{\partial \mathcal{L}}{\partial x} = \left[y^T A = w^T \right] =$$

$$= \frac{\partial \mathcal{L}}{\partial x}\left[w^T x \right] = w^T$$

$$\mathcal{L} = \mathcal{L}^T = x^T A^T y$$

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial \mathcal{L}^T}{\partial y} = x^T A^T$$

$$\underset{1\times 1}{\mathcal{L}} = \underset{1\times n}{x^T} \underset{n\times n}{A} \underset{n\times 1}{x} = \sum_{j=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j$$

$$\frac{\partial \mathcal{L}}{\partial x} - ?$$

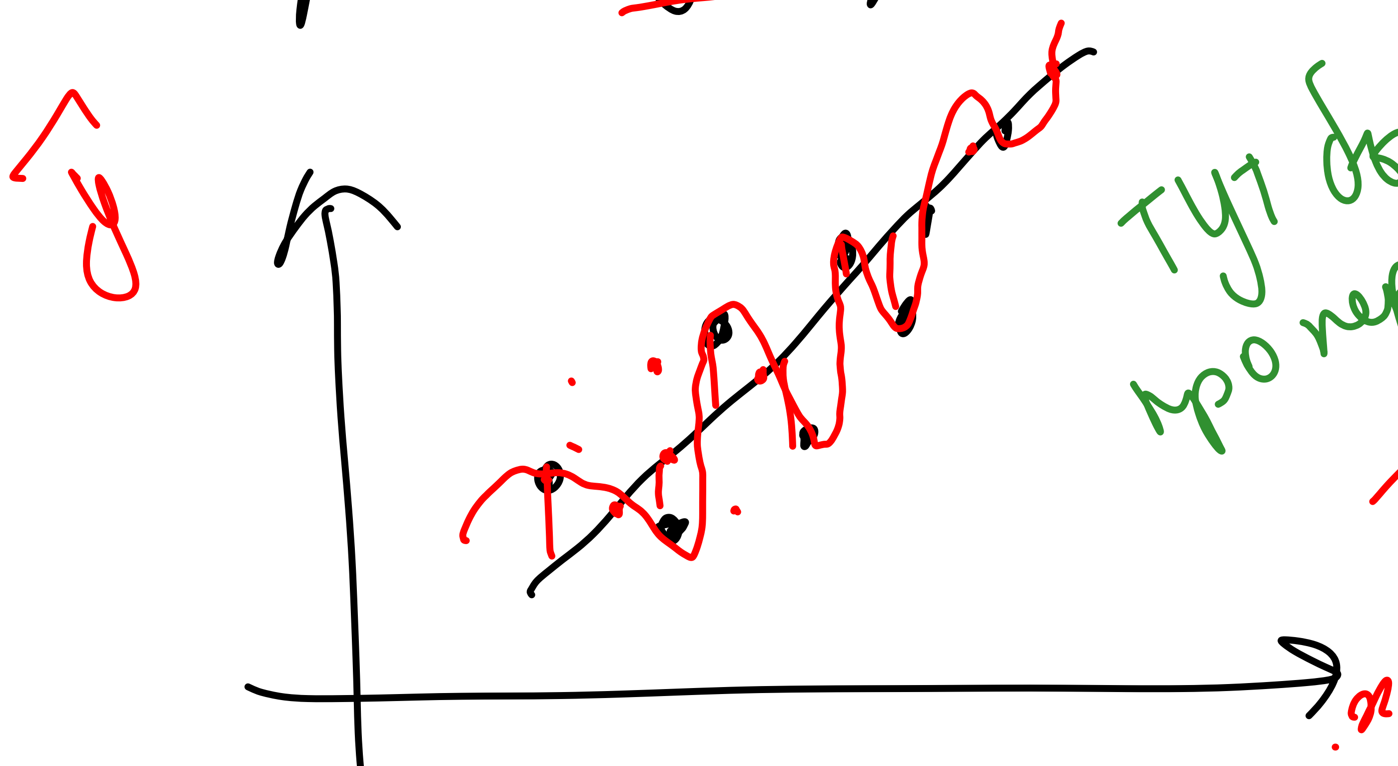
$$\frac{\partial \mathcal{L}}{\partial x_k} = \overbrace{\sum_{j=1}^{n} \mathcal{L}_{kj} x_j} + \overbrace{\sum_{i=1}^{n} a_{ik} x_i}$$

$$\frac{\partial \mathcal{L}}{\partial x} = x^T A^T + x^T A = \boxed{x^T (A^T + A)}$$

A - сим. матр.

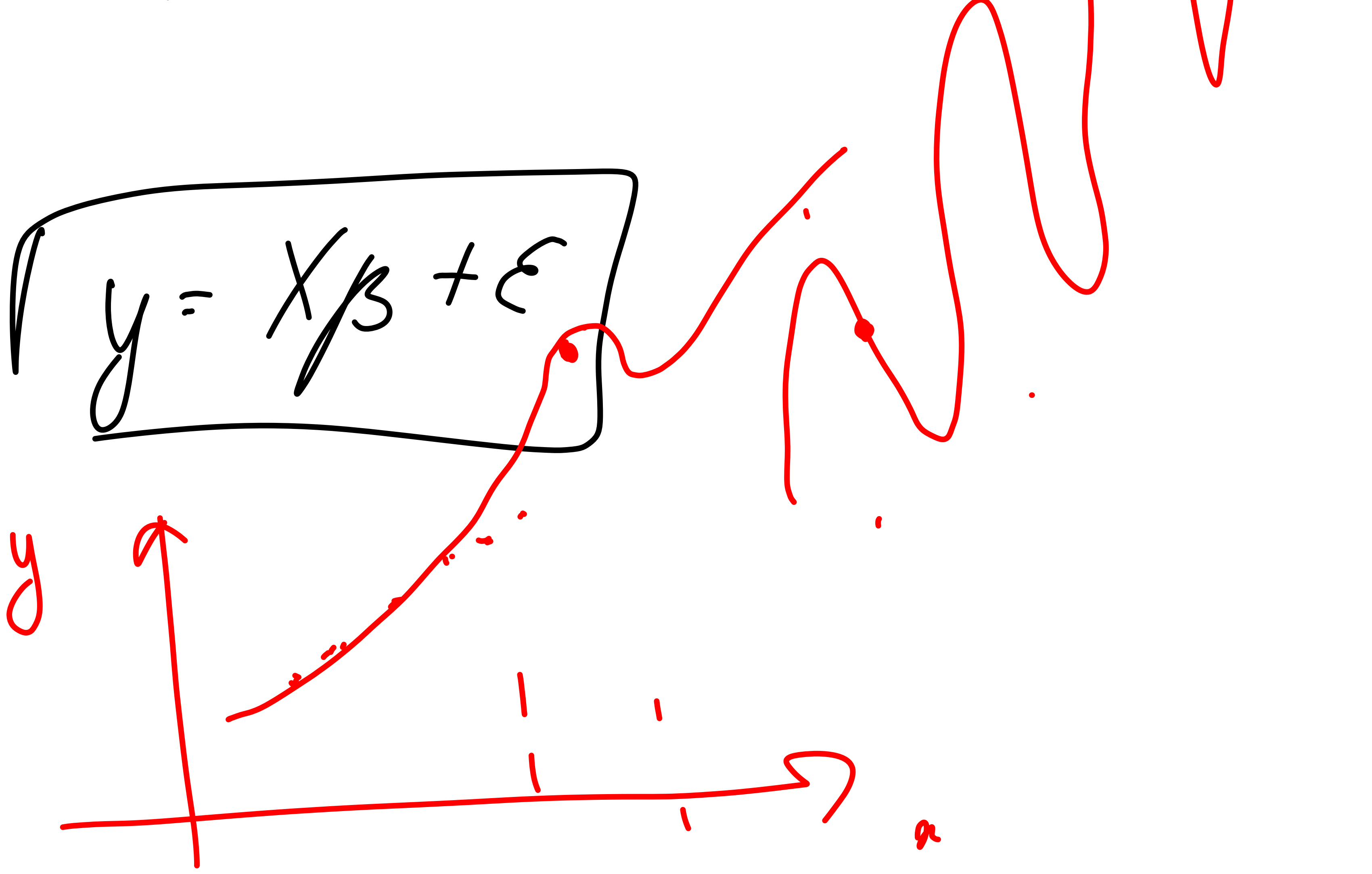
$$\frac{\partial L}{\partial x} = 2x^T A$$

$$Q(\beta) = (y - X\beta)^T(y - X\beta) \to \min_{\beta}$$



тут было переобучение

$$\boxed{y = X\beta + \varepsilon}$$



$$\beta = (\beta_1, \beta_2 \ \dots \ \beta_m)$$ Ridge-регр-ия

$$Q(\beta) = (\underset{n\times 1}{y} - \underset{n\times m}{X}\underset{m\times 1}{\beta})^T (y - X\beta)_{n\times 1} + \lambda^{\geq 0} \beta^T \beta \to \min$$

$$d(AB) = A\,dB + dA \cdot B$$

$$\frac{\partial Q(\beta)}{\partial \beta} = -\underbrace{X^T_{m\times n}(y - X\beta)_{n\times 1}}_{\boxed{m\times 1}} - (y - X\beta)^T_{1\times n} \underset{n\times m}{X} + 2\lambda\beta_{m\times 1} =$$

Вопрос на подумать: а так вообще законно?

$$= -2X^T(y - X\beta) + 2\lambda\beta$$

$$\frac{\partial((y - X\beta)^T(y - X\beta))}{\partial \beta} =$$

$$= \frac{\partial(y - X\beta)^T}{\partial \beta}(y - X\beta) + (y - X\beta)^T \frac{\partial(y - X\beta)}{\partial \beta} =$$

$$-2X^T(y - X\hat{\beta}_R) = -2\lambda\hat{\beta}_R$$

$$-X^Ty + X^TX\hat{\beta}_R + \lambda\hat{\beta}_R = 0.$$

$$(\overset{n\times n}{X^TX} + \overset{n\times m}{\lambda I})\hat{\beta}_R = X^Ty$$

$$\hat{\beta}_R = (X^TX + \lambda I)^{-1}X^Ty$$

$\lambda = 0 \longrightarrow$ МНК

$\lambda \to \infty \quad \hat{\beta}_R \to 0$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$E\left((y - \hat{y})^2\right)$$

? $\cancel{\sum_{i=1}^{n}(y_i - \hat{y}_i)}$

MAE

$$\frac{1}{n}\sum_i |y_i - \hat{y}_i| \;,\quad E|y_{\cancel{i}} - \hat{y}_{\cancel{i}}|$$

$E|Y-c|$, m - медиана

показали, что лучшее предск. для Y, которое мин. MAE - это медиана

~~Му сп~~

$a \neq m$

$E|Y-m| \leq E|Y-a|$

сл. величина

$E(|Y-a|-|Y-m|) \geq 0$

$\boxed{m < a}$ аналогично для $m > a$

$\underline{Y \leq m}$

$|Y-a|-|Y-m| = a-Y-(m-Y) = a-m$

$\underline{Y > m}$

$|Y-a|-|Y-m| \geq Y-a-(Y-m) = m-a$.

$Z = |Y-a|-|Y-m|$

$I = [Y \leq m]$

$E(Z) = E(ZI) + E(Z(1-I)) =$

$$\geq (a-m)E[I] + (m-a)E(1-I) =$$

$$= (a-m)P(Y \leq m) + (m-a)P(Y > m) =$$

$$= (a-m)(\underbrace{2P(Y \leq m) - 1}) \geq 0$$

$$P(Y \leq m) \geq \frac{1}{2}$$

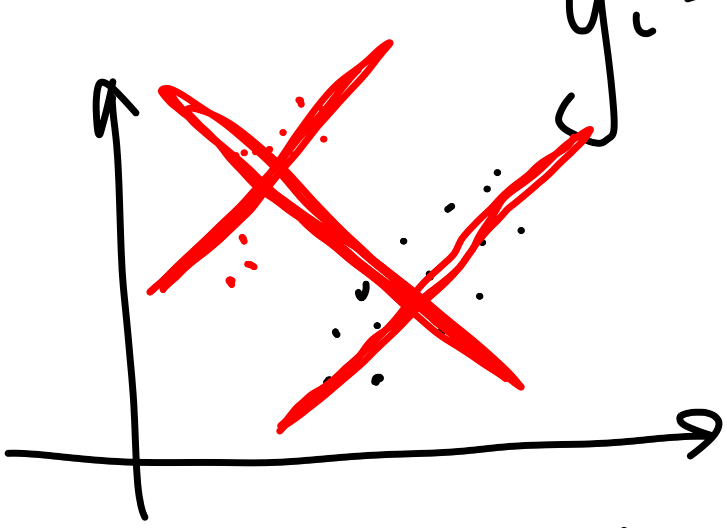
(т.к. m – медиана)

$$P(Y \geq m) \geq \frac{1}{2}$$

$\Rightarrow$ m минимизирует MAE

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ – одна регрессия

$\hat{y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 \bar{x}_i$ – вторая



$\hat{\beta}_1 > 0$

$\hat{\alpha}_1 > 0$

$\hat{z}_i = \hat{\gamma}_0 + \hat{\gamma}_1 \ddot{x}_i$ – все вместе