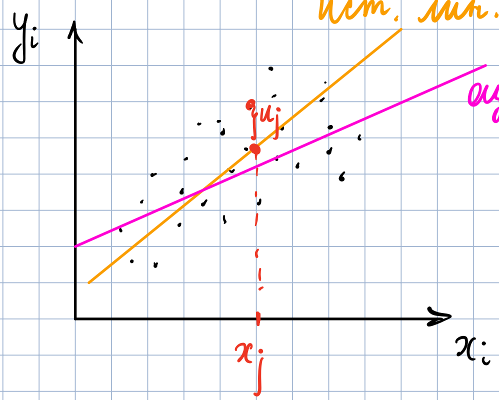
$$\boxed{y_i = w_0 + w_1 X_{1i} + w_2 X_{2i} + \dots + w_k X_{ki}} + u_i$$

цел — y_i

коэф / веса — $w_0, w_1, w_2, \dots, w_k$

признаки рег. — $X_{1i}, X_{2i}, \dots, X_{ki}$

сл. ошиб. — u_i



$\hat{w}_0$ — оценка w_0, получ. по выб-е

$$\boxed{\hat{y}_i = \hat{w}_0 + \hat{w}_1 X_{1i} + \dots + \hat{w}_k X_{ki}}$$

оцен. линия регрессии

$\hat{w}_0, \dots \hat{w}_k$ — рез-т миним. ф. потерь

→ MSE: $\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2 \to \min\limits_{\hat{w}_0, \dots \hat{w}_k}$ — классика

$$\frac{1}{n} \sum_i (y_i - \hat{w}_0 - \hat{w}_1 X_{1i} - \dots - \hat{w}_k X_{ki}) \to \min\limits_{\hat{w}_0 \dots \hat{w}_k}$$

→ MAE: $\frac{1}{n} \sum_i |y_i - \hat{y}_i| \to \min\limits_{\hat{w}_0 \dots \hat{w}_k}$ — медиан. регрессия

Виды

парная

$y_i = w_0 1 + w_1 X_{1i} + u_i$

$\hat{y}_i = \hat{w}_0 + \hat{w}_1 X_{1i}$

множ-я

$y_i = w_0 + w_1 X_{1i} + \dots + w_k X_{ki} + u_i$

$\underset{n \times 1}{y} = \underset{n \times (k+1)}{X} \underset{(k+1) \times 1}{w} + \underset{n \times 1}{u}$

Ф. потерь

MSE → кв.

MAE → мед.

Huber → ..

...

Метрики

MSE, MAE, R^2;

MSE → ЛР1( ), ЛР2( ) — MSE, MAE, R^2

MAE → ЛР1

MSE → ЛР2

— MSE, MAE, R^2

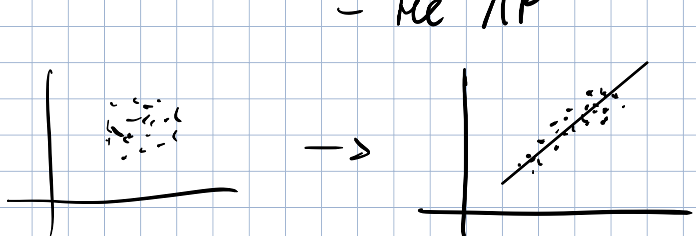
По каким перем. ЛР линейна?

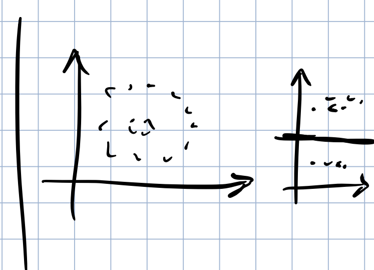
$$y_i = w_0 + w_1 x_{1i} + \ldots + w_k x_{ki} + u_i$$

$$y_i = w_0 + w_1 \ln x_{1i} + \ldots + w_k x_{ki}^2 + u_i \text{ — ЛР}$$

$$\ln y_i = w_0 + w_1 x_{1i} \cdot x_{2i} + \ldots + w_k e^{x_{ki}} + \ln u_i \text{ — ЛР}$$

$$y_i = w_0 + \ln w_1 \cdot x_{1i} + \ldots + w_k^2 x_{ki} + u_i \text{ — не ЛР}$$



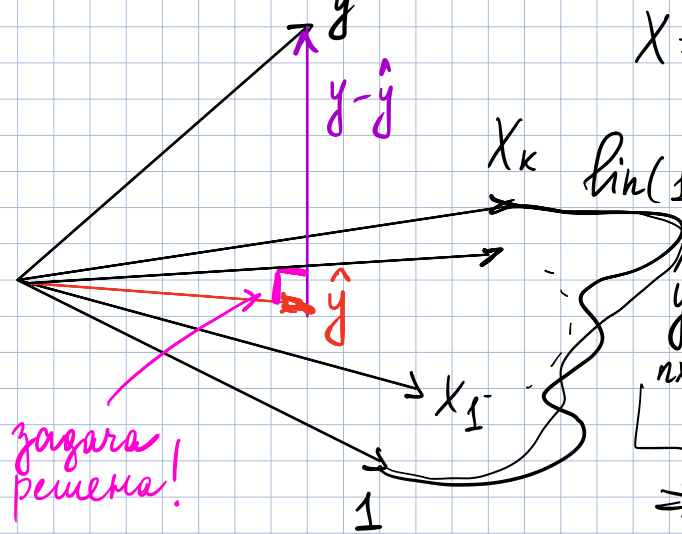


① Клас. вариант:

$$\underset{n\times 1}{y} = \underset{n\times(k+1)}{X}\,\underset{(k+1)\times 1}{w} + u_{n\times 1} \qquad MSE \to \min_{\hat w}$$

$$\boxed{\hat y = X\hat w} \qquad \boxed{\hat w = (X^TX)^{-1}X^Ty}$$

$(n \gg k)$



$$X = \begin{bmatrix} | & | & | & & | \\ 1 & x_1 & x_2 & \dots & x_k \\ | & | & | & & | \end{bmatrix}_{n\times(k+1)}$$

$$\underbrace{\underset{n\times 1}{\hat y} = \hat w_0 + \hat w_1 \underset{n\times 1}{x_1} + \dots + \hat w_k \underset{n\times 1}{x_k}}$$

$$\Rightarrow \hat y \in lin(1, x_1, \dots x_k)$$

$$MSE = \frac{1}{n}\sum_i (y_i - \hat y_i)^2 = \boxed{\frac{1}{n}\|y - \hat y\|_2^2}$$

$$\hat y = X\hat w$$

$$y - \hat y \perp 1, x_1 \dots x_k$$

$$\langle y - \hat y, x_j \rangle = 0$$

$$x_j^T(y - \hat y) = 0 \quad \forall j \Rightarrow X^T(y - \hat y) = 0$$

$$X^T(y - X\hat w) = 0$$

$$X^Ty = X^TX\hat w$$

$$\hat w = (X^TX)^{-1}X^Ty$$

Мультиколлинеарность

теор.

в X есть ЛЗ призн.

$(X^T X)^{-1}$ не сущ-т

$\hat{w}$ не сущ-т

$X_j = a_1 X_m + a_2 X_p + \ldots + a_* X_*$

$X_1 \quad X_2 \quad X_1 + X_2$

1

⋮

1000

$(X_1 + X_2) = X_1 + X_2$

прак.

призн. в X сильно коррел.

$\hat{w}$ существ.

$\hat{w}$ численно неуст.

$\hat{w}$ имеют большую дисперсию

1. Легко идент.

$\boxed{corr(X)}$

cond numb(X)

2. Не особо большие проблемы

3. Легко бороться

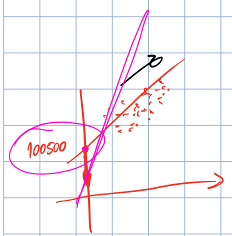
а) Выбросить призн.

б) М/к из-за X_j ; $X_j := X_j + norm(0, 10^{-3})$

в) Регуляр.

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

$$\hat{w}^* = (X^T X + \lambda I)^{-1} X^T y \qquad \lambda \geq 0$$

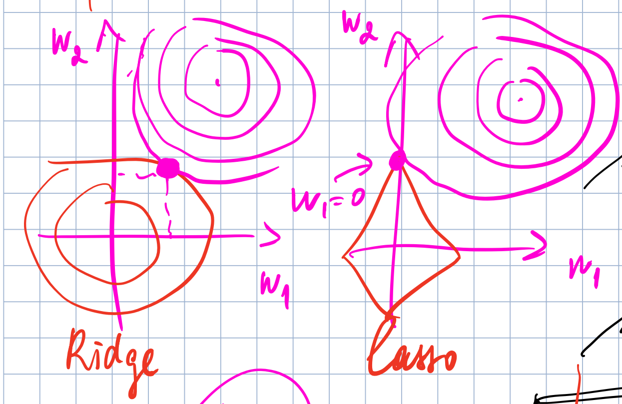


$$MSE^* = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2 + \underbrace{\lambda \sqrt{\sum_{k>0} w_k^2}}_{} \to \min_{\hat{w}_0, \dots, \hat{w}_k}$$

$\hat{w}_0$

Ridge

$$+\lambda \sum_{k>0} |w_k| \quad \text{— lasso}$$



$$+\underbrace{\cancel{\lambda}\left(a_1 \sqrt{\sum_{k>0} w_k^2} + a_2 \sum_{k>0} |w_k|\right)}_{\text{elastic net}}$$

1	$X_1 \dots X_k$	z^1	z^2	z^3	sum (z_1, z_2, z_3)	$1 = z^1 + z^2 + z^3$
1	…	1	0	0	1	0
1	…	1	0	0	1	0
1	…	0	1	0	1	1
⋮		…	…	…	⋮	…
1	…	0	0	1	1	1

dummy trap

R^2

Задача из тетрадки

$MSE \to \min$

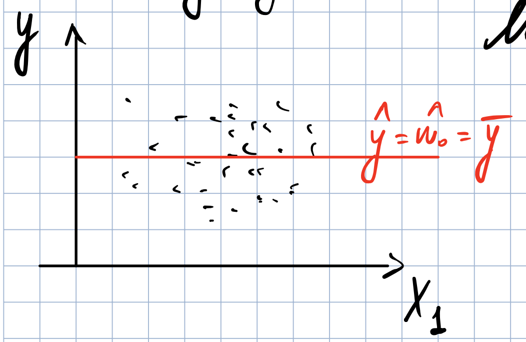
$\hat{y} = C$

$y_i = w_0 + u_i$

$\hat{y}_i = \hat{w}_0$

$\hat{y} = C = w_0$ — регрессия на конст.

$\hat{w}_0 = \hat{y} = \bar{y}$



$$MSE = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2 \to \min_{\hat{w}_0}$$

$$= \frac{1}{n} \sum_i (y_i - \hat{w}_0)^2 \to \min_{\hat{w}_0}$$

$$MSE'_{\hat{w}_0} = \cancel{\frac{1}{n}(-1)} \cdot 2 \sum_i (y_i - \hat{w}_0) = 0$$

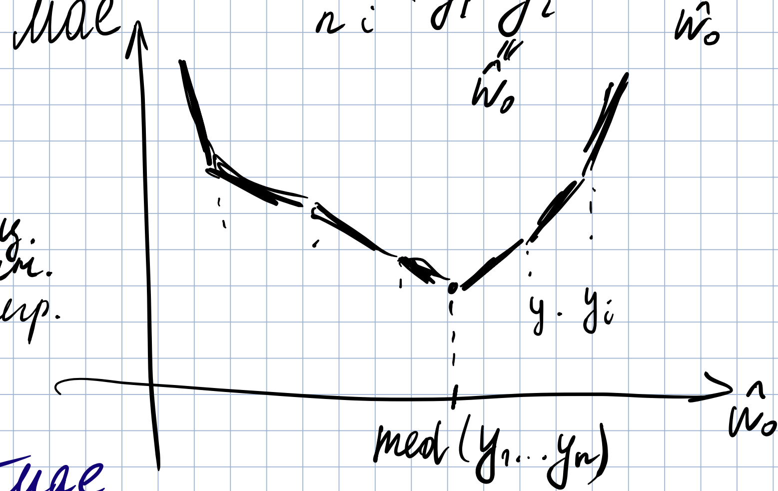
$$\sum_i y_i = \underbrace{\sum_i \hat{w}_0}_{= n \cdot \hat{w}_0}$$

$$\hat{w}_0 = \frac{\sum_i y_i}{n} = \bar{y}$$

MSE → min :

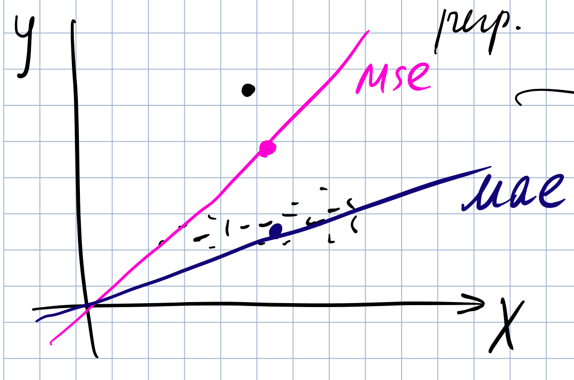
$\left(\frac{\sum x_i}{n}, \frac{\sum_i y_i}{n}\right) \in$ ол. лин. рег.

$$MAE : \frac{1}{n} \sum_i |y_i - \underbrace{\hat{y}_i}_{= \hat{w}_0}| \to \min_{\hat{w}_0}$$



mae → min :

(med(X), med(y)) $\in$ ол. лин. рег.



$$\bar{y} := \frac{\sum_i y_i}{n} \qquad \boxed{\bar{y} = \bar{\hat{y}}}$$

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2 / n-1}{\sum_i (y_i - \bar{y})^2 / n-1} = \frac{\text{выб. дисп. прогноза}}{\text{выб. дисп. } y}$$

$\boxed{R^2 \in [0;1] \text{ – статист.}}$

может

МО: $R^2 \notin [0;1]$

без $\hat{w}_0$

др. ф. потерь