HA 3 Bubnov Dmitriy gr.4

№1

$$X = \begin{pmatrix} x_1 & x_1 \\ x_2 & x_2 \\ \vdots & \vdots \\ x_n & x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \hat{y} = \hat{\beta}_1 X + \hat{\beta}_2 X = \quad \to \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$$

a) $$loss(\hat{\beta}) = (y-\hat{y})^T (y-\hat{y}) + \lambda \cdot \hat{\beta}^T \hat{\beta}$$

$$loss(\hat{\beta}) = (y-\hat{y})^T (y-\hat{y}) + \lambda \hat{\beta}^T \hat{\beta} \to \min_{\hat{\beta}}$$

$$d\left((y-\hat{y})^T (y-\hat{y}) + \lambda \hat{\beta}^T \hat{\beta}\right) = d\left((y-\hat{y})^T (y-\hat{y})\right) + d(\lambda \hat{\beta}^T \hat{\beta}) =$$

$$= 2(y-\hat{y})^T d(y-\hat{y}) + 2\lambda \hat{\beta}^T d\hat{\beta} = -2(y-\hat{y})^T d\hat{y} + 2\lambda \hat{\beta}^T d\hat{\beta} =$$

$$= -2(y-\hat{y})^T d(\hat{\beta}_1 X + \hat{\beta}_2 X) + 2\lambda \hat{\beta}^T d\hat{\beta}) =$$

$$= -2(y-\hat{y})^T (\underset{0}{dX}\hat{\beta} + X d\hat{\beta}) + 2\lambda \hat{\beta}^T d\hat{\beta} =$$

$$= -2(y-\hat{y})^T \cdot X d\hat{\beta} + 2\lambda \hat{\beta}^T d\hat{\beta} = 0$$

$$-\cancel{2} (y - X\hat{\beta})^T \cdot X \cancel{d\hat{\beta}} + \cancel{2}\lambda \hat{\beta}^T \cancel{d\hat{\beta}} = 0$$

$$-(y - X\hat{\beta})^T \cdot X + \lambda \hat{\beta}^T = 0$$

$$(-y - X\hat{\beta}^T X + \lambda \hat{\beta}^T)^T = 0$$

$$-X^T(y - X\hat\beta) + \lambda\hat\beta = 0$$

$$-X^T y + X^T X\hat\beta + \lambda\hat\beta = 0$$

$$X^T X\hat\beta + \lambda\hat\beta = X^T y$$

$$\left[\ \hat\beta = (X^T X + \lambda I)^{-1} X^T y\ \right]$$

$$X^T X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_1 & x_2 & & x_n \end{pmatrix} \cdot \begin{pmatrix} x_1 & x_1 \\ \vdots & \vdots \\ x_n & x_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

$$X^T X + \lambda I = \begin{pmatrix} \sum x_i^2 + \lambda & \sum x_i^2 \\ \sum x_i^2 & \sum x_i^2 + \lambda \end{pmatrix}$$

$$(X^T X + \lambda I)^{-1} = \frac{1}{(\sum x_i^2 + \lambda)^2 - (\sum x_i^2)^2} \cdot \begin{pmatrix} \sum x_i^2 + \lambda & -\sum x_i^2 \\ -\sum x_i^2 & \sum x_i^2 + \lambda \end{pmatrix}$$

$$X^T y = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum x_i y_i \end{pmatrix}$$

$$\hat\beta = \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} = \frac{1}{(\sum x_i^2 + \lambda)^2 - (\sum x_i^2)^2} \cdot \begin{pmatrix} \sum x_i^2 + \lambda & -\sum x_i^2 \\ -\sum x_i^2 & \sum x_i^2 + \lambda \end{pmatrix} \cdot \begin{pmatrix} \sum x_i y_i \\ \sum x_i y_i \end{pmatrix} =$$

$$= \begin{pmatrix} \dfrac{\lambda \sum x_i y_i}{(\sum x_i^2 + \lambda)^2 - (\sum x_i^2)^2} \\ \dfrac{\lambda \sum x_i y_i}{(\sum x_i^2 + \lambda)^2 - (\sum x_i^2)^2} \end{pmatrix} \Rightarrow \hat\beta_1 = \hat\beta_2 = \frac{\cancel{\lambda} \sum x_i y_i}{\cancel{(\sum x_i^2)^2} + 2\cancel{\lambda} \sum x_i^2 + \lambda^2 - \cancel{(\sum x_i^2)^2}} =$$

$$= \frac{\sum x_i y_i}{2\sum x_i^2 + \lambda} \qquad \left[\ \hat\beta_1 = \hat\beta_2 = \frac{\sum x_i y_i}{2\sum x_i^2 + \lambda}\ \right]$$

в) $\lambda \to \infty$

$$\hat{\beta}_i = \lim_{\lambda \to \infty} \frac{\sum x_i y_i}{2\sum x_i^2 + \lambda} = 0$$

c)

$$\hat{\beta}_1 + \hat{\beta}_2 = \lim_{\lambda \to 0} \frac{\sum x_i y_i}{2\sum x_i^2 + \lambda} + \lim_{\lambda \to 0} \frac{\sum x_i y_i}{2\sum x_i^2 + \lambda} = \lim_{\lambda \to 0} \frac{2\sum x_i y_i}{2\sum x_i^2 + \lambda} =$$

$$= \frac{\sum x_i y_i}{\sum x_i^2}$$

№2

$$y = X\beta + u \qquad E(u|X=0) \qquad \underset{[n \times k]}{X} = \left(\quad \right)$$

$$rk(X) = k \qquad Var(u|X) = \sigma^2 W \qquad W \neq I$$

$$\hat{\beta} = \hat{\beta}^{OLS}$$

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

$$E(\hat{\beta}|X) = E((X^T X)^{-1} X^T y | X) = (X^T X)^{-1} X^T E(X\beta + u|X) =$$

$$= \cancel{(X^T X)^{-1}} \cdot \cancel{X^T X}\beta + 0 = \beta$$

$$E(\hat{\beta}) = E(E(\hat{\beta}|X)) = \beta$$

б) $Var(\hat{\beta}|X) = Var((X^T X)^{-1} X^T y | X) =$

$$= (X^T X)^{-1} X^T \underbrace{Var(X\beta + u|X)}_{= Var(u|X) = \sigma^2 W} \cdot ((X^T X)^{-1} X^T)^{-1} =$$

$= (X^T X)^{-1} X^T \cdot \sigma^2 \cdot W \left((X^T X)^{-1} X^T\right)^{-1}$

c) $W \neq I$ means that there is heteroscedasticity in our case $\Rightarrow$ CI won't be valid

d) $Cov(y, \hat{\beta} | X) = Cov(y, (X^T X)^{-1} X y | X) =$

$= (X^T X^{-1}) X \underbrace{Cov(y, y | X)}_{= Var(y|X) = Var\, u = \sigma^2 W} = (X^T X^{-1}) X \sigma^2 W$

№3

a) $X^T X = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}$

$\det \begin{pmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{pmatrix} = 0 \Rightarrow 36 - 12\lambda + \lambda^2 - 1 = 0$

$\Downarrow$

$\lambda_1 = 7 \qquad \lambda_2 = 5$

$\lambda = 7$:

$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \tilde{v}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \quad C = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$

$\lambda = 5$: $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \tilde{v}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$

$D = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix}$

$X^T X = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$

b)

$$X = U \cdot D \cdot V^T \qquad U^T U = I$$

$$V^T V = I$$

c)

№ 4

d) $\begin{cases} \|Xw\|^2 \to \max\limits_w \\ \text{s.t.} \\ \|w\|^2 = 1 \end{cases}$ $\begin{cases} (Xw)^T Xw \to \max\limits_w \\ \text{s.t.} \\ w^T w = 1 \end{cases}$

$$\begin{cases} w^T X^T \cdot Xw \to \max \\ \text{s.t.} \\ w^T w = 1 \end{cases}$$

$$L = w^T X^T X w - \lambda(w^T w - 1)$$

e) F.O.C.

$$dL(w) = dw^T X^T Xw + w^T d(X^T Xw) - 2\lambda w^T dw =$$

$$= dw^T X^T Xw + w^T X^T X dw - 2\lambda w^T dw = 2w^T X^T X dw -$$

$$- 2\lambda w^T dw = 2w^T (X^T X - \lambda I) dw$$

f) $2w^T (X^T X - \lambda I) = 0$

$(X^T X - \lambda I)^T w = 0$ λ – eigenvalue of $X^T X$

w – eigenvector of $X^T X$

$$w_1^* = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \qquad w_2^* = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

$$\|Xw_1^*\| = w_1^{*T} X^T X w_1^* = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = 7$$

$$\|Xw_2^*\| = w_2^{*T} X^T X w_2^* = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} = 5$$

$7 > 5 \Rightarrow$ optimal w is $w^* = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$