HA 3

Kataeva Anna
gr 1

1 $loss(\hat\beta) = (y-\hat y)^T(y-\hat y) + \lambda\hat\beta^T\hat\beta, \quad \hat y = \hat\beta_1 x + \hat\beta_2 x$

a) find optimal $\hat\beta_1, \hat\beta_2$

$$X = \begin{pmatrix} x_1 & x_1 \\ x_2 & x_2 \\ \vdots & \vdots \\ x_n & x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \hat y = \begin{pmatrix} \hat y_1 \\ \hat y_2 \\ \vdots \\ \hat y_n \end{pmatrix} \quad \hat\beta = \begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} \quad \begin{pmatrix} \hat y_1 \\ \hat y_2 \\ \vdots \\ \hat y_n \end{pmatrix} = \begin{pmatrix} x_1 & x_1 \\ x_2 & x_2 \\ \vdots & \vdots \\ x_n & x_n \end{pmatrix}\begin{pmatrix} \hat\beta_1 \\ \hat\beta_2 \end{pmatrix} = \begin{pmatrix} \hat\beta_1 x_1 + \hat\beta_2 x_1 \\ \hat\beta_1 x_2 + \hat\beta_2 x_2 \\ \vdots \\ \hat\beta_1 x_n + \hat\beta_2 x_n \end{pmatrix}$$

$$\min_{\hat\beta} loss(\hat\beta) = \min_{\hat\beta}\left[(y-\hat y)^T(y-\hat y) + \lambda\hat\beta^T\hat\beta\right]$$

$$d\left[(y-\hat y)^T(y-\hat y) + \lambda\hat\beta^T\hat\beta\right] = 2(y-\hat y)^T d(y-\hat y) + 2\lambda\hat\beta^T d\hat\beta = -2(y-\hat y)^T d\hat y + 2\lambda\hat\beta^T d\hat\beta =$$

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

$$= -2(y-X\hat\beta)^T\cdot X\,d\hat\beta + 2\lambda\hat\beta^T d\hat\beta = \left[-2(y-X\hat\beta)^T\cdot X + 2\lambda\hat\beta^T\right]d\hat\beta$$

F.O.C.

$$\underbrace{-\cancel{2}(y-X\hat\beta)^T\cdot X}_{[1\times 2]} + \underbrace{\cancel{2}\lambda\hat\beta^T}_{[1\times 2]} = \underset{[2\times 1]}{0}$$

$$\left[-(y-X\hat\beta)^T X + \lambda\hat\beta^T\right]^T = 0^T$$

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

$$-X^Ty + X^TX\hat\beta + \lambda\hat\beta = 0$$

$$X^TX\hat\beta + \lambda\hat\beta = X^Ty$$

$$(X^TX + \lambda I)\hat\beta = X^Ty$$

$$\underbrace{\hat\beta}_{2\times 1} = (\underbrace{X^TX}_{2\times 2} + \lambda\underbrace{I}_{2\times 2})^{-1}\underbrace{X^Ty}_{2\times 1} \qquad \boxed{\hat\beta = (X^TX + \lambda I)^{-1}X^Ty}$$

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

$$X^TX + \lambda I = \begin{pmatrix} \sum_{i=1}^n x_i^2 + \lambda & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^2 + \lambda \end{pmatrix}$$

$$(X^TX + \lambda I)^{-1} = \frac{1}{\left(\sum_{i=1}^n x_i^2 + \lambda\right)^2 - \left(\sum_{i=1}^n x_i^2\right)^2}\cdot\begin{pmatrix} \sum_{i=1}^n x_i^2 + \lambda & -\sum_{i=1}^n x_i^2 \\ -\sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^2 + \lambda \end{pmatrix}$$

$$X^T y = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum x_i y_i \end{pmatrix}$$

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

$$= \begin{pmatrix} \dfrac{\sum x_i y_i \cdot \lambda}{(\sum x_i^2 + \lambda)^2 - (\sum x_i^2)^2} \\[2ex] \dfrac{\sum x_i y_i \cdot \lambda}{(\sum x_i^2 + \lambda)^2 - (\sum x_i^2)^2} \end{pmatrix}$$

$$\cancel{(\sum x_i^2)^2} + 2\lambda \sum x_i^2 + \lambda^2 - \cancel{(\sum x_i^2)^2} = 2\lambda \sum x_i^2 + \lambda^2$$

$$\hat\beta_2 = \hat\beta_1 = \frac{\sum x_i y_i \cdot \lambda}{(\sum x_i^2 + \lambda)^2 - (\sum x_i^2)^2} = \frac{\sum x_i y_i \cdot \lambda}{\cancel{(\sum x_i^2)^2} + 2\sum x_i^2 \cdot \lambda + \lambda^2 - \cancel{(\sum x_i^2)^2}} = \frac{\sum x_i y_i \cdot \cancel{\lambda}}{\cancel{\lambda}(2\sum x_i^2 + \lambda)} = \boxed{\frac{\sum x_i y_i}{2\sum x_i^2 + \lambda}}$$

b) $\lambda \to \infty$

$$\lim_{\lambda\to\infty} \hat\beta_1 = \lim_{\lambda\to\infty} \hat\beta_2 = \lim_{\lambda\to\infty} \left(\frac{\sum x_i y_i}{2\sum x_i^2 + \lambda} \right) = \lim_{\lambda\to\infty} \left(\frac{\frac{\sum x_i y_i}{\lambda} \to 0}{\frac{2\sum x_i^2}{\lambda} \to 0 + 1} \right) = 0$$

c) $$\lim_{\lambda\to 0} (\hat\beta_1 + \hat\beta_2) = \lim_{\lambda\to 0} \left[\frac{2\sum x_i y_i}{2\sum x_i^2 + \cancel{\lambda}\to 0} \right] = \frac{\cancel{2}\sum x_i y_i}{\cancel{2}\sum x_i^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

2. Consider the model $y = X\beta + u$ where β is non-random, $\mathbb{E}(u \mid X) = 0$, the matrix X of size $n \times k$ has rank $X = k$, but $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 W$ with $W \neq I$. Let $\hat\beta$ be the standard OLS estimator of β.

(a) Find $\mathbb{E}(\hat\beta \mid X)$, $\mathbb{E}(\hat\beta)$.

(b) Find $\mathbb{V}\mathrm{ar}(\hat\beta \mid X)$.

(c) How do you think, will the standard confidence interval for β be valid in this case?

(d) Find $\mathbb{C}\mathrm{ov}(y, \hat\beta \mid X)$.

$y = X\beta + u$, β - non-random

$E(u|X) = 0$ $\quad \underset{[n\times k]}{X}$: rank $X = k$, $Var(u|X) = \sigma^2 W$, $W \neq I$

a) $E(\hat\beta | X)$ - ? $\quad E(\hat\beta)$ - ?

OLS estimator $\hat\beta = (X^T X)^{-1} \cdot X^T y$

$$E(\hat\beta | X) = E\left[(X^T X)^{-1} X^T y \mid X\right] = (X^T X)^{-1} X^T E(X\beta + u \mid X) =$$

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

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

b) $$Var(\hat\beta | X) = Var\left[(X^T X)^{-1} X^T y \mid X\right] = (X^T X)^{-1} X^T Var(X\beta + u | X) \left[(X^T X)^{-1} X^T\right]^T \ominus$$

$$Var(X\beta + u | X) = Var(u|X) = \sigma^2 W$$

$$\ominus\ (X^T X)^{-1} X^T \sigma^2 W X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1} X^T W X (X^T X)^{-1}$$

c) When $W = I$, we have : $Var(\hat\beta | X) = \sigma^2 (X^T X)^{-1} \cancel{X^T X} \cancel{(X^T X)^{-1}} =$

$$= \sigma^2 (X^T X)^{-1}$$

and for this case $CI_{1-\alpha}(\beta)$ is valid : $CI_{1-\alpha}(\beta) = \hat\beta \pm t_{\alpha/2}(n-k)\, \sigma \sqrt{(X^T X)^{-1}}$

However, in our case $Var(\hat\beta | X) = \sigma^2 (X^T X)^{-1} X^T W X (X^T X)^{-1}$. Thus, heteroscedasticity may be presented $\Rightarrow$ t-test are invalid $\Rightarrow$ $CI_{1-\alpha}(\beta)$ will be also invalid.

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

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

$$Cov(y, y | X) = Var(y|X) = Var(X\beta + u | X) = Var(u|X) = \sigma^2 W$$

3

$$X = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}$$

a)

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

$$\det(X^T X - \lambda I) = \begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = (6-\lambda)^2 - 1 = 36 - 12\lambda + \lambda^2 - 1 = \lambda^2 - 12\lambda + 35 = 0$$

$\lambda_1 = 5 \quad \lambda_2 = 7$

$\underline{\lambda_1 = 5}$:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \to (1 \quad 1) \qquad v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad \tilde{v}_1 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$\underline{\lambda_2 = 7}$:

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \to (-1 \quad 1) \qquad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \tilde{v}_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$$P = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \qquad P^{-1} = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

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

b) SVD of X - ?

$X = UDV^T$, where $U^T U = I\ (3\times 3)$

$V^T V = I\ (2 \times 2)$

X - 3×2 $\qquad$ D - diagonal (3×2)

$$X^T X = (UDV^T)^T UDV^T = VD^T \underbrace{U^T U}_{I} DV^T = VD^T D V^T = \underbrace{V \begin{pmatrix} d_{11}^2 & 0 \\ 0 & d_{22}^2 \end{pmatrix} V^T}_{\Downarrow}$$

$D^T D$:

$$\begin{pmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & 0 \end{pmatrix} \begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} d_{11}^2 & 0 \\ 0 & d_{22}^2 \end{pmatrix}$$

$$V = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \qquad d_{11} = \sqrt{7} \qquad d_{22} = \sqrt{5}$$

$$XX^T = UDV^T \cdot (UDV^T)^T = UD \underbrace{V^T \cdot V}_{I} D^T U^T = UDD^T U^T$$

DD^T:

$$\begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & 0 \end{pmatrix} = \begin{pmatrix} d_{11}^2 & 0 & 0 \\ 0 & d_{22}^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$XX^T = \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}$$

$$\det(XX^T - \lambda I) = \begin{vmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{vmatrix} = (5-\lambda)^2(2-\lambda) - \left(9(5-\lambda) + 5-\lambda\right) = -\lambda(5-\lambda)(7-\lambda) = 0$$

$$\lambda_1 = 0 \quad \lambda_2 = 5 \quad \lambda_3 = 7$$

$\lambda_1 = 0$:

$$\begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \rightarrow V_1 = \begin{pmatrix} -3 \\ -1 \\ 5 \end{pmatrix} \qquad \tilde{V}_1 = \begin{pmatrix} -3/\sqrt{35} \\ -1/\sqrt{35} \\ 5/\sqrt{35} \end{pmatrix}$$

$\lambda_2 = 5$:

$$\begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \longrightarrow V_2 = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} \qquad \tilde{V}_2 = \begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \\ 0 \end{pmatrix}$$

$\lambda_3 = 7$:

$$\begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \rightarrow V_3 = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \qquad \tilde{V}_3 = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix}$$

$$U = \begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} & -3/\sqrt{35} \\ 1/\sqrt{14} & 3/\sqrt{10} & -1/\sqrt{35} \\ 2/\sqrt{14} & 0 & 5/\sqrt{35} \end{pmatrix}$$

SVD of X:

$$\boxed{X = \begin{pmatrix} 3/\sqrt{14} & -1/\sqrt{10} & -3/\sqrt{35} \\ 1/\sqrt{14} & 3/\sqrt{10} & -1/\sqrt{35} \\ 2/\sqrt{14} & 0 & 5/\sqrt{35} \end{pmatrix}\begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^T}$$

c)

$$\hat{X} = u_1 \cdot d_{11} \cdot V_1^T = \begin{pmatrix} 3/\sqrt{14} \\ 1/\sqrt{14} \\ 2/\sqrt{14} \end{pmatrix} \cdot \sqrt{7} \cdot \left(-\frac{1}{\sqrt{2}} \;\; \frac{1}{\sqrt{2}}\right)^T = \begin{pmatrix} 3/\sqrt{2} \\ 1/\sqrt{2} \\ 2/\sqrt{2} \end{pmatrix}\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \boxed{\begin{pmatrix} -\frac{3}{2} & \frac{3}{2} \\ -\frac{1}{2} & \frac{1}{2} \\ -1 & 1 \end{pmatrix}}$$

4 $X = UDV^T$

$d_{11} > d_{22} > \dots > 0$

maximize $\|Xw\|^2$ by choosing optimal w s.t. $\|w\|^2 = 1$

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

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

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

e) $\underline{\text{F.O.C.}}$

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

$d(w^T X^T X w) = dw^T \cdot X^T X w + w^T d(X^T X w) = dw^T X^T X w + w^T \cdot X^T X dw = 2w^T X^T X dw$

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

$[(X^T X - \lambda I)]^T w = 0$

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

from b): $X^T X = V D^T D V^T$

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

Thus, $w_1^* = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ or $w_2^* = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ $\|w_1^*\| = \|w_2^*\| = 1$ (constraint is satisfied)

Let's check:

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

V

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

Thus $\boxed{w_{max} = w_1^* = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}}$