Chemerisov Artem HA3

Problem 1.

$\hat{y} = \hat{\beta}_1 X + \hat{\beta}_2 X$, $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$, $X = \begin{pmatrix} x \\ x \end{pmatrix}$

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

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

$$d\, loss(\hat{\beta}) = d\left[(y-\hat{y})^T(y-\hat{y})\right] + d(\lambda \hat{\beta}^T \cdot \hat{\beta}) = -2(y - X\hat{\beta})^T \cdot X d\hat{\beta} + \overbrace{\lambda \hat{\beta}^T d\hat{\beta} + \lambda d\hat{\beta}^T \hat{\beta}}^{2\lambda d\hat{\beta} \hat{\beta}^T} = 0$$

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

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

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

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

$$\hat{\beta}\lambda = X^T y - X^T X \hat{\beta}$$

$\hat{\beta} = (\lambda + X^T X)^{-1} \cdot X^T y$ - optimal $\hat{\beta}_1$ and $\hat{\beta}_2$ for fixed λ

b) $\lim_{\lambda \to \infty} \hat{\beta} = \frac{X^T \cdot y}{(\lambda + X^T X)} = \frac{X^T y}{\infty} = 0$, $(\hat{\beta}_1, \hat{\beta}_2) = (0, 0)$

c) $\lim_{\lambda \to 0}(\hat{\beta}_1 + \hat{\beta}_2) = \lim_{\lambda \to 0}(\lambda + X^T X)^{-1} = (X^T X)^{-1} X^T y = \hat{\beta}^{ols}$

Problem 2.

$y = X\beta + u$, $E(u|X)=0$, $Var(u|X) = \sigma^2 W$ $\hat{\beta} = \hat{\beta}^{OLS}$

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

$= (X^T X)^{-1} \cdot (X^T X)\beta + \overset{\nearrow 0}{E(u|X)} = \beta$

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

b) $Var(\hat{\beta}|X) = Var((X^T X)^{-1} X^T y | X) = ((X^T X)^{-1} X^T) Var(X\overset{\nearrow const}{\beta} + u | X)((X^T X)^{-1} X^T)^T =$

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

c) In this case the standard confidence interval would be invalid as $Var(u|X) \neq \sigma^2$, thus there is a non constant variance that leads to heteroscedacity, CI will be biased and inconsistent.

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

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

Problem 3.

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

a) $X^T = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$ $\quad X^T X = \underset{[2\times 3]}{\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}} \underset{[3\times 2)}{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}} = \underset{[2\times 2]}{\begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}} = A$

$$\det(A - \lambda I) = \begin{vmatrix} 6-\lambda & 1 \\ 1 & 6-\lambda \end{vmatrix} = (6-\lambda)^2 = 6^2 - 12\lambda + \lambda^2 \Rightarrow \lambda_1 = 5 \quad \lambda_2 = 7$$

$$D = \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix} \quad v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad D = P^{-1} A P = \begin{pmatrix} 5 & 0 \\ 0 & 7 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

b) $X \cdot X^T = \underset{[3\times 2)}{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}} \underset{[2\times 3]}{\begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}} = \underset{[3\times 3]}{\begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix}}$

$$\begin{pmatrix} 5-\lambda & 0 & 3 \\ 0 & 5-\lambda & 1 \\ 3 & 1 & 2-\lambda \end{pmatrix} = (2-\lambda)(5-\lambda)^2 - (9(5-\lambda) + (5-\lambda)) = (5-\lambda)((5-\lambda)(2-\lambda) - 10) = (5-\lambda)(-5\lambda - 2\lambda + \lambda^2)$$

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

$\lambda_1 = 5$:

$X^T X \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad v_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad v_1^n = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

$X X^T \quad \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 1 \\ 3 & 1 & -3 \end{pmatrix} \quad v_1' = \begin{pmatrix} \frac{1}{3} \\ -1 \\ 0 \end{pmatrix} \quad v_1'^n = \begin{pmatrix} \frac{1}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} \\ 0 \end{pmatrix}$

$\lambda_2 = 7$:

$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad v_2^n = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

$\begin{pmatrix} -2 & 0 & 3 \\ 0 & -2 & 1 \\ 3 & 1 & -5 \end{pmatrix} \quad v_2' = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \quad v_2'^n = \begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix}$

$\lambda_3 = 0$:

$X X^T \quad \begin{pmatrix} 5 & 0 & 3 \\ 0 & 5 & 1 \\ 3 & 1 & 2 \end{pmatrix} \quad v_3 = \begin{pmatrix} 3 \\ 1 \\ -5 \end{pmatrix} \quad v_3^n = \begin{pmatrix} \frac{3}{\sqrt{35}} \\ \frac{1}{\sqrt{35}} \\ -\frac{5}{\sqrt{35}} \end{pmatrix}$

$$X = U D V^T = \underset{[3\times 3]}{\begin{pmatrix} \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{35}} \\ \frac{1}{\sqrt{14}} & -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{35}} \\ \frac{2}{\sqrt{14}} & 0 & -\frac{5}{\sqrt{35}} \end{pmatrix}} \underset{[3\times 2]}{\begin{pmatrix} \sqrt{7} & 0 \\ 0 & \sqrt{5} \\ 0 & 0 \end{pmatrix}} \underset{[2\times 2]}{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}} = \underset{[3\times 2]}{\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 1 & 1 \end{pmatrix}}$$

c) $\begin{pmatrix} \frac{3}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix} \cdot (\sqrt{7}) \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 1 \end{pmatrix} \rightarrow (1\ 1)$ matrix approximation of X with rank $= 1$

$U_{1^{st} \text{ column}}$ $\quad D_{1^{st} \text{ column}}$ $\quad V^T_{1^{st} \text{ row}}$

Problem 4

d)

$$\begin{cases} \|X_w^2\|^2 \to \max\limits_w \\ \text{s.t.} \\ \|w\|^2 = 1 \end{cases} \qquad \|X_w\|^2 = w^T X^T X w \qquad \|w\|^2 = w^T w \quad \Rightarrow \begin{cases} w^T X^T X w \to \max\limits_w \\ \text{s.t.} \\ w^T w = 1 \end{cases}$$

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

e) FOC:

$$\begin{cases} L'_w = (dw)^T X^T X w + w^T X^T X\, dw - \lambda (dw)^T w - \lambda w^T dw = 0 \\ L'_\lambda = -w^T w + 1 = 0 \end{cases}$$

$$\begin{cases} L'_w = 2w^T (X^T X - \lambda I) = 0 \\ L'_\lambda = 1 - w^T w = 0 \end{cases} \qquad \begin{cases} (X^T X - \lambda I) w = 0 \\ w^T w = 1 \end{cases}$$

f) $W = V^T w$

$w = VW$

$(X^T X - \lambda I) w\, dw = 0$

$d(UDW)^T (UDW) - \lambda\, d(VW) = 0$

$2W^T D^T U^T U D\, dW - \lambda V dW = 0$

$2W^T dW = \lambda V dW$

$W^T = \frac{\lambda V}{2}$

$V^T w = \frac{(\lambda V)^T}{2}$

$w = \frac{\lambda}{2} \cdot V^T \cdot (V^T)^{-1}$