HW1 _ Yuting - Sun

P1

a). Proof. To prove X^TX is positive definite,

pick a vector $\vec{Y}$ (Dimension - $p \times 1$)

$$\vec{Y}^T (X^T X) \vec{Y} = (X\vec{Y})^T X \vec{Y} = \|X\vec{Y}\|_2^2$$

$$= \left\| \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{n1} & \cdots & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} \right\|_2^2$$

$$= \left\| \begin{bmatrix} \sum_{j=1}^{p} x_{1j} y_j \\ \sum_{j=1}^{p} x_{2j} y_j \\ \vdots \\ \sum_{j=1}^{p} x_{nj} y_i \end{bmatrix} \right\|_2^2 = \left(\sum_{j=1}^{p} x_{1j} y_j\right)^2 + \cdots \left(\sum_{j=1}^{p} x_{nj} y_j\right)^2.$$

$\underset{\substack{(n\times p)\\ n>p}}{X}$ - full rank matrix. $\Rightarrow$ for each row $\underbrace{x_{i1}, x_{i2} \cdots x_{ip}}$ $i \in [1,n]$ are linearly independent.

$\Rightarrow M = \sum_{j=1}^{p} x_{i1} y_1 = x_{i1} y_1 + x_{i2} y_2 + \cdots x_{ip} y_p = 0$, only when $y_1 = \cdots = y_p = 0$.

Since $\vec{Y}$ is not $\vec{0}$, hence, at least one element (y_i) in $\vec{Y}$ is not 0. $\Rightarrow M = \sum_{j=1}^{p} x_{ij} y_j \neq 0$.

Therefore, $\vec{Y}^T (X^T X) \vec{Y} = \left\| \begin{bmatrix} \sum_{j=1}^{p} x_{ij} Y_i \\ \vdots \\ \sum_{j=1}^{p} x_{nj} y_j \end{bmatrix} \right\|_2^2 > 0.$

X^TX is positive definite.

b). Proof :

To prove XX^T is positive semi-definite,

Pick a vector $\vec{R}$ (Dimension: $n\times 1$)

$$\vec{R}^T X X^T \vec{R} = (X^T\vec{R})^T X^T \vec{R} = \|X^T\vec{R}\|_2^2$$

$$= \left\| \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ x_{1p} & \cdots & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} \sum_{i=1}^n x_{i1} r_i \\ \sum_{i=1}^n x_{i2} r_i \\ \vdots \\ \sum_{i=1}^n x_{ip} r_i \end{bmatrix} \right\|_2^2 = \left(\sum_{i=1}^n x_{i1} r_i\right)^2 + \cdots \left(\sum_{i=1}^n x_{ip} r_i\right)^2.$$

$\because$ $n > p$, for X^T, it results in that $x_{1j}\ x_{2j} \cdots x_{nj}$ $j\in[1,p]$ are linearly dependent.

$\therefore$ By picking suitable coefficients ($r_1 - r_n$), we can make:

$$N = \sum_{i=1}^n x_{i1} r_i = x_{11}r_1 + x_{12}r_2 + \cdots x_{in}r_n = 0 \quad (\exists\, r_i,\ i\in[1,p])$$

Therefore, $\vec{R}^T X X^T \vec{R} = \left\| \begin{bmatrix} \sum_{i=1}^n x_{i1} r_i \\ \vdots \\ \sum_{i=1}^n x_{ip} r_i \end{bmatrix} \right\|^2 \geqslant 0$

X^TX is a square $(p \times p)$ matrix with p linearly independent eigenvectors. X^TX can be factorized as. $X^TX = Q \Lambda Q^T$.

$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}$ Q – orthogonal eigenvector matrix $\quad \underset{p\times n}{Q}\ \underset{n\times n}{\Lambda}\ \underset{n\times p}{Q^T}$.

Pick vector. $\vec{q}_i (p\times 1)$ – column vectors in matrix Q
$i = 1, 2 \dots n$.

$\vec{q}_i^{\,T} X^TX \vec{q}_i = \underline{\vec{q}_i^{\,T} Q} \Lambda \underline{Q^T \vec{q}_i} = \sum_{i=1}^{n} \lambda_i \underline{q_i^T q_i}\, q_i^T q_i$ ③

$= \underbrace{(Q^T \vec{q}_i)^T \Lambda (Q^T \vec{q}_i)}_{\text{scaler.}}$

① $q_i^T q_i = 1 \neq 0$ ($\because$ q_i is eigenvector.)

Since. ② $\vec{q}_i^{\,T} X^TX \vec{q}_i > 0$ (X^TX is positive definite)

therefore, based on ①②③. $\lambda_i > 0$. $\quad i = 1, 2 \dots n$.

$\underbrace{tr(X^TX)}_{\text{diagonal.}} = tr(Q\Lambda Q^T) = tr(Q^TQ\Lambda) = tr(I\Lambda) = tr(\Lambda)$

$= \sum_{i=1}^{n} \lambda_i > 0.$
$(\lambda_i > 0)$

P2. Proof.

To minimize the noise sourse,

$$\hat{\theta}_{LS} = \arg\min_{\theta} \frac{1}{2} \| \vec{y} - X\vec{\theta} \|_2^2 .$$

$$f(\vec{\theta}_{LS}) = \frac{1}{2} \| \vec{y} - X\vec{\theta} \|_2^2$$

$$= \frac{1}{2} (\vec{y} - X\vec{\theta})^T (\vec{y} - X\vec{\theta})$$

$$= \frac{1}{2} \left(\vec{y}^T\vec{y} - \vec{y}^T X\vec{\theta} - (X\vec{\theta})^T\vec{y} + (X\vec{\theta})^T (X\vec{\theta}) \right)$$

$$\frac{d f(\vec{\theta}_{LS})}{d\vec{\theta}} = -\frac{1}{2} \left(2\vec{X}^T (\vec{y} - X\vec{\theta}) \right) = -X^T\vec{y} + X^T X\vec{\theta} .$$

$$\frac{d^2 f(\vec{\theta}_{LS})}{d\vec{\theta}^2} = X^T X \quad (PD) \Rightarrow \text{increasing function.}$$

Using Newton's iterative minimization method,

$$\vec{\theta}^{(i)} = \vec{\theta}^{(i-1)} - \mu_i (X^T X)^{-1} (-X^T\vec{y} + X^T X \vec{\theta}^{(i-1)})$$

$\uparrow$ step length

$$= \vec{\theta}^{(i-1)} - \mu_i (X^T X)^{-1} (-X^T\vec{y}) - \mu_i \underbrace{(X^T X)^{-1} X^T X}_{I} \vec{\theta}^{(i-1)}$$

if $\mu_i = 1$ $$= \vec{\theta}^{(i-1)} + (X^T X)^{-1} (X^T\vec{y}) - \vec{\theta}^{(i-1)} = (X^T X)^{-1} (X^T\vec{y})$$

$$\frac{\partial f(\vec{\theta}^{(i)})}{\partial \vec{\theta}^{(i)}} = -X^T\vec{y} + X^T X\vec{\theta}^{(i)} = -X^T\vec{y} + \underbrace{X^T X (X^T X)^{-1}}_{I} (X^T\vec{y})$$

$$= -X^T\vec{y} + X^T\vec{y} = \vec{0} .$$

$\Rightarrow$ $\theta^{(i)} = \theta^{(i-1)}$ $\Rightarrow$ Newton's method can converge in one step.

*

P 3.

a). $y = X\beta + \varepsilon$

Suppose there are 3 teams and they make up 5 match-ups,

consider the linear model; $y_i^H - y_i^A = \theta_{H_i} - \theta_{A_i} + \mu + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$

e.g. $y_1^H - y_1^A = \theta_{H_1} - \theta_{A_1} + \mu + \varepsilon_1$

$y_2^H - y_2^A = \theta_{H_2} - \theta_{A_2} + \mu + \varepsilon_2$

$\vdots$

$y_5^H - y_5^A = \theta_{H_5} - \theta_{A_5} + \mu + \varepsilon_5$

$$\beta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \mu \end{bmatrix}$$

$$\Rightarrow \vec{y} = \underbrace{\begin{bmatrix} 1 & -1 & 0 & 1 \\ 1 & 0 & -1 & 1 \\ 0 & 1 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}}_{X} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

$$N(X) = \begin{bmatrix} 1 & -1 & 0 & 1 \\ 1 & 0 & -1 & 1 \\ 0 & 1 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ -1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} a-b+d=0 \\ a-c+d=0 \\ b-c+d=0 \\ -a+b+d=0 \\ -a+c+d=0 \end{cases}$$

$a-b+d=0,\ a-c+d=0 \Rightarrow b=c.$

$b-c+d=0,\ -a+b+d=0 \Rightarrow a=b$

$-a+c+d=0 \rightarrow a=c \quad d=0$

$\Rightarrow a=b=c$, $d=0$

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} a \\ a \\ a \\ 0 \end{bmatrix} = \underset{\text{const}}{a} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \Rightarrow \text{span}(x) = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

$$y_i^H - y_i^A = \overbrace{\begin{pmatrix} 0 \cdots 0 & \underset{\theta_{Hi}}{1} & -0- & \underset{\theta_{Ai}}{-1} & 0 \cdots 0 & \underset{\mu}{1} \end{pmatrix}}^{X_i} \vec{\beta} + \varepsilon_i$$

$$\underset{31\times 1}{\beta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{30} \\ \mu \end{bmatrix}$$

$$\vec{y} = \underbrace{\begin{pmatrix} - & x_1 & - \\ & \vdots & \\ - & x_i & - \\ & \vdots & \\ - & x_{1408} & - \end{pmatrix}}_{X} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{30} \\ \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_{1408} \end{bmatrix}$$

For each row in X,

$$\begin{array}{c} \begin{array}{ccccc} i & & j & & \end{array} \\ \begin{bmatrix} 1 & 0 \cdots & -1 & 0 \cdots & 1 \\ 1 & -1 & 0 & 0 \; 0 \cdots & 1 \\ & h & \vdots & & \\ 0 & 1 & -1 & 0 \cdots & 1 \end{bmatrix} \\ \begin{array}{ccc} & & k \end{array} \end{array} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{31} \end{bmatrix} = \vec{0}$$

$$\begin{aligned} &m_i - m_j + m_{31} = 0 \\ &m_i - m_h + m_{31} = 0 \\ &\quad\vdots \\ &m_h - m_k + m_{31} = 0 \\ &\quad\vdots \end{aligned} \quad \Rightarrow \quad \begin{aligned} &m_j = m_h = \cdots = m_k = m \\ &m_{31} = 0 \end{aligned}$$

$$\Rightarrow \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{31} \end{bmatrix} = \begin{bmatrix} m \\ m \\ \vdots \\ m \\ 0 \end{bmatrix} = \underset{\substack{\uparrow \\ \text{const}}}{m} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}.$$

$$\Rightarrow \vec{\gamma} = (1, 1, \cdots 1, 0)^T.$$

b). As proved before,

$$f(\hat{\beta}) = \frac{1}{2}(\vec{y} - X\vec{\beta})^T(\vec{y} - X\vec{\beta})$$

$$\frac{df(\hat{\beta})}{d\vec{\beta}} = -X^T(\vec{y} - X\vec{\beta}) = 0 \qquad \left(\frac{d^2 f(\hat{\beta})}{d\vec{\beta}^2} = 2X^TX > 0\right.$$

$$X^T\vec{y} = X^TX\vec{\beta} \qquad \vec{\beta} = (X^TX)^{-1}X^T\vec{y}.$$

Since, null space of X is the span of the vector $\begin{bmatrix}1\\ \vdots\\ 0\end{bmatrix}$

$$X \cdot \begin{bmatrix}1\\ \vdots\\ 0\end{bmatrix} t = \vec{0}$$

if we pick $\beta' = \hat{\beta} + \overbrace{\begin{bmatrix}1\\ \vdots\\ 0\end{bmatrix}}^{\gamma} t$ which is different from β.

then, $X\beta' = X(\hat{\beta} + \underbrace{\begin{bmatrix}1\\ \vdots\\ 0\end{bmatrix}}_{\gamma} t) = X\hat{\beta} + \underbrace{X\begin{bmatrix}1\\ \vdots\\ 0\end{bmatrix} t}_{0} = X\hat{\beta}$

Since $f(\hat{\beta}) = \frac{1}{2}\|\vec{y} - X\hat{\beta}\|_2^2$ is the minimized function

$(\beta' \neq \hat{\beta})$ $f(\beta') = \frac{1}{2}\|\vec{y} - \underbrace{X\beta'}_{= X\hat{\beta}}\|_2^2 = \frac{1}{2}\|\hat{y} - X\hat{\beta}\|_2^2 = f(\hat{\beta})$ is also the minimized function.

Therefore, the solution is not unique.

c) $L(\vec{\beta}, \vec{\lambda}) = \frac{1}{2}\|\vec{y} - X\vec{\beta}\|_2^2 + \lambda(\vec{r}^T\vec{\beta})$

$$\frac{\partial L}{\partial \vec{\beta}} = -X^T(\vec{y} - X\vec{\beta}) + \lambda\vec{r}. = \vec{0}$$

$$\frac{\partial L}{\partial \vec{\lambda}} = \vec{r}^T\vec{\beta} = 0.$$

$$\begin{bmatrix} X^TX & \vec{r} \\ \vec{r}^T & 0 \end{bmatrix}\begin{bmatrix} \vec{\beta} \\ \lambda \end{bmatrix} = \begin{bmatrix} X^T\vec{y} \\ 0 \end{bmatrix}$$

Assume there are two different solutions $\vec{\beta}_1, \vec{\beta}_2, \vec{\lambda}_1, \vec{\lambda}_2$,

$\vec{\beta}_1 \neq \vec{\beta}$ $\quad \vec{\lambda}_1 \neq \vec{\lambda}_2$.

$$\begin{bmatrix} X^TX & \vec{r} \\ \vec{r}^T & 0 \end{bmatrix}\begin{bmatrix} \vec{\beta}_1 - \vec{\beta}_2 \\ \lambda_1 - \lambda_2 \end{bmatrix} = \begin{bmatrix} \vec{0} \\ 0 \end{bmatrix}$$

$$\begin{cases} X^TX(\vec{\beta}_1 - \vec{\beta}_2) + \vec{r}(\lambda_1 - \lambda_2) = \vec{0} & (1) \\ \vec{r}^T \cdot (\vec{\beta}_1 - \vec{\beta}_2) = 0. & (2) \end{cases}$$

For (1) $\quad X^TX(\vec{\beta}_1 - \vec{\beta}_2) + \vec{r}(\vec{\lambda}_1 - \vec{\lambda}_2) = \vec{0}.$

Multiply by $(\vec{\beta}_1 - \vec{\beta}_2)^T$,

$$(\vec{\beta}_1 - \vec{\beta}_2)^T(X^TX)(\vec{\beta}_1 - \vec{\beta}_2) + \overset{(3)}{(\vec{\beta}_1 - \vec{\beta}_2)^T\vec{r}(\vec{\lambda}_1 - \vec{\lambda}_2)} = 0$$

Since given (2). $\vec{r}^T \cdot (\vec{\beta}_1 - \vec{\beta}_2) = (\vec{\beta}_1 - \vec{\beta}_2)^T\vec{r} = 0$,

$(3) = \vec{0}. \Rightarrow (\vec{\beta}_1 - \vec{\beta}_2)^T X^TX(\vec{\beta}_1 - \vec{\beta}_2) = \|X(\vec{\beta}_1 - \vec{\beta}_2)\|_2^2 = 0.$

$\Rightarrow \vec{\beta}_1 = \vec{\beta}_2.$ contradicts with our assumption.

Therefore, the sollution is unique.

(7)

d)

① $\nabla_{\vec{\beta}} L(\vec{\beta}, \lambda) = 0$

$\Rightarrow \quad -X^T(\vec{y} - X\vec{\beta}) + \lambda\vec{r} = 0$

$$\underline{X^T X\vec{\beta} + \lambda\vec{r} = X^T\vec{y}} \quad ③$$

② $\nabla_{\lambda} L(\vec{\beta}, \lambda) = 0$

$$\underline{\vec{r}^T\vec{\beta} = 0} \quad ④$$

Based on ③④.

$$\left[\begin{array}{c|c} X^T X & \vec{r} \\ \hline \vec{r}^T & 0 \end{array}\right] \left[\begin{array}{c} \vec{\beta} \\ \hline \lambda \end{array}\right] = \left[\begin{array}{c} X^T\vec{y} \\ \hline 0 \end{array}\right].$$

has the same form as. $\left[\begin{array}{c|c} \mathbb{I} & b^T \\ \hline b & 0 \end{array}\right] \left[\begin{array}{c} \vec{\beta} \\ \hline \lambda \end{array}\right] = \left[\begin{array}{c} \vec{z} \\ \hline 0 \end{array}\right]!$

P 4

a) Proof: sample mean $\Rightarrow \frac{1}{n}\sum_{i=1}^{n} y_i$

$$E\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(y_i) = \frac{1}{n}\cdot\left(E(y_1) + E(y_2) + \cdots E(y_n)\right)$$

$\because \{y_i\}_{i=1}^{n}$ iid $E(y_1) = E(y_2) = \cdots E(y_n) = \mu$,

$\therefore E\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\cdot n\mu = \mu. \Rightarrow$ sample mean is unbiased.

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(y_i).$$

$(y_i)_{i=1}^{n}$ iid $Var(y_1) = \cdots Var(y_n) = \sigma^2$.

$$\Rightarrow Var\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0.$$

b). Proof. Since $(y_i)_{i=1}^{n}$ iid with bounded variance.

hence, $P(|\bar{y}_n - \mu| > \varepsilon) \le \frac{Var(\bar{y}_n)}{\varepsilon^2}$ (Chebyshev's inequality)

$= \frac{\sigma^2}{n\varepsilon^2}$ which tends to 0 as $n\to\infty$.

Therefore sample mean convergos to the true mean in probability.

P5

proof:

$$\hat{\theta}_b = (1+\alpha)\hat{\theta}_u.$$

$$MSE(\hat{\theta}_b) = E[(\hat{\theta}_b - \theta_0)^2]$$

$$= E[((\hat{\theta}_b - E(\hat{\theta}_b)) + (E(\hat{\theta}_b) - \theta_0))^2]$$

$$= E[(\hat{\theta}_b - E(\hat{\theta}_b))^2 + \underbrace{(E(\hat{\theta}_b) - \theta_0)^2}_{\text{const}}$$

$$+ 2(\hat{\theta}_b - E(\hat{\theta}_b))(E(\hat{\theta}_b) - \theta_0)]$$

$$= E[(\hat{\theta}_b - E(\hat{\theta}_b))^2] + (E(\hat{\theta}_b) - \theta_0)^2$$

$$+ 2E[(\hat{\theta}_b - E(\hat{\theta}_b))]\underbrace{(E(\hat{\theta}_b) - \theta_0)}_{\text{const}=0}$$

$$= E[(\hat{\theta}_b - E(\hat{\theta}_b))^2] + (E(\hat{\theta}_b) - \theta_0)^2.$$

Since $\hat{\theta}_b = (1+\alpha)\hat{\theta}_u$, $MSE(\hat{\theta}_u) = E[(\hat{\theta}_u - E(\hat{\theta}_u))^2] + (E(\hat{\theta}_u) - \theta_0)^2$

$$MSE(\hat{\theta}_b) = (1+\alpha)^2 E[(\hat{\theta}_u - E(\hat{\theta}_u))^2] + [(1+\alpha)E(\hat{\theta}_u) - \theta_0]^2$$

$$= (1+\alpha)^2 E[(\hat{\theta}_u - E(\hat{\theta}_u))^2] + (1+\alpha)^2(E(\hat{\theta}_u) - \theta_0)^2$$

$$\underline{-(1+\alpha)^2\theta_0^2} + 2(1+\alpha)^2\theta_0 E(\hat{\theta}_u) \underline{+\theta_0^2} - 2\theta_0(1+\alpha)E(\hat{\theta}_u)$$

$$= (1+\alpha)^2 \{E[(\hat{\theta}_u - E(\hat{\theta}_u))^2 + [E(\hat{\theta}_u) - \theta_0)^2]$$

$$-(\alpha^2+2\alpha)\theta_0^2 + 2\theta_0[(1+\alpha)^2 - (1+\alpha)] \cdot E(\hat{\theta}_u)$$

$$= (1+\alpha)^2 MSE(\hat{\theta}_u) - (\alpha^2+2\alpha)\theta_0^2 + 2(\alpha^2+\alpha)\underline{\theta_0 E(\hat{\theta}_u)}$$

θ_0^2

$E(\hat{\theta}_u) = \theta_0$

$$= (1+\alpha)^2 MSE(\hat{\theta}_u) + \alpha^2 \cdot \theta_0^2.$$

$$\because MSE(\hat{\theta}_b) < MSE(\hat{\theta}_u)$$

$$(1+\alpha)^2 MSE(\hat{\theta}_u) + \alpha^2\theta_0^2 < MSE(\hat{\theta}_u)$$

$\because \alpha < 0 \quad (\alpha^2 + 2\alpha) MSE(\hat{\theta}_u) + \alpha^2 \theta_0^2 < 0$

$$\alpha(\alpha+2) MSE(\hat{\theta}_u) + \alpha \cdot \alpha \theta_0^2 < 0$$

$$(\alpha+2) MSE(\hat{\theta}_u) + \alpha \theta_0^2 > 0$$

$$\alpha \cdot (MSE(\hat{\theta}_u) + \theta_0^2) > -2 MSE(\hat{\theta}_u)$$

$$\alpha > \frac{-2 MSE(\hat{\theta}_u)}{MSE(\hat{\theta}_u) + \theta_0^2}.$$

On the other hand,

$$-\left(\frac{2 MSE(\hat{\theta}_u)}{MSE(\hat{\theta}_u) + \theta_0^2}\right) > -\left(\frac{2 \cancel{MSE(\hat{\theta}_u)}}{\cancel{MSE(\hat{\theta}_u)}}\right) = -2.$$

Therefore. $-2 < -\dfrac{2 MSE(\hat{\theta}_u)}{MSE(\hat{\theta}_u) + \theta_0^2} < \alpha < 0$

P6. $\quad y = X\theta + \varepsilon$

$\because \theta \sim N(\theta_0, \Sigma_0) \quad \epsilon \sim N(0, \Sigma_\epsilon) \qquad \theta \perp\!\!\!\perp \varepsilon.$

$$y \sim N(X\theta_0, X\Sigma_0 X^T + \Sigma_\epsilon)$$

$$Cov(\theta, y) = Cov(\theta, X\theta + \varepsilon) = \Sigma_0 X^T.$$

$$Cov(y, \theta) = Cov(X\theta + \varepsilon, \theta) = X\Sigma_0$$

For Gaussian distribution,

the joint distribution of θ, y : $\begin{pmatrix}\theta \\ y\end{pmatrix} \sim N\left(\begin{pmatrix}\theta_0 \\ X\theta_0\end{pmatrix}, \begin{pmatrix}\Sigma_0 & \Sigma_0 X^T \\ X\Sigma_0 & X\Sigma_0 X^T + \Sigma_\epsilon\end{pmatrix}\right)$

$$\theta | y \sim N(\theta_0 + \Sigma_0 X^T (X\Sigma_0 X^T + \Sigma_\epsilon)^{-1}(y - X\theta_0), \Sigma_0 - \Sigma_0 X^T (X\Sigma_0 X^T + \Sigma_\epsilon) X\Sigma_0)$$

$$\Rightarrow E(\theta|y) = \theta_0 + \Sigma_0 X^T (X\Sigma_0 X^T + \Sigma_\epsilon)^{-1}(y - X\theta_0)$$

$$Var(\theta|y) = \Sigma_0 - \Sigma_0 X^T (X\Sigma_0 X^T + \Sigma_\epsilon)^{-1} X\Sigma_0$$