Hw 2 6648-5170-90 Yu Wei Chen

1.

Least Squared Estimate : $\underline{\hat{\beta}} = (X^TX)^{-1}X^T\underline{y}$

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

$E[\hat{\theta}] = E[\underline{a}^T(X^TX)^{-1}X^T\underline{y}]$

$= \underline{a}^T(X^TX)^{-1}X^T E[\underline{y}]$ $\quad \underline{y} = X\underline{\beta} + \epsilon$, $E[\epsilon] = 0$

$= \underline{a}^T\underline{\beta}$

$\tilde{\theta} = \underline{a}^T(X^TX)^{-1}X^T\underline{y} + \underline{d}^T\underline{y} = \underline{c}^T\underline{y}$

$\underline{c}^T = \underline{a}^T(X^TX)^{-1}X^T + \underline{d}^T$, $\underline{y} = X\underline{\beta} + \varepsilon$

$E[\tilde{\theta}] = E[\underline{a}^T(X^TX)^{-1}X^T(X\underline{\beta}) + \underline{a}^T(X^TX)^{-1}X^T\varepsilon + \underline{d}^TX\underline{\beta} + \underline{d}^T\varepsilon]$

$= E[\underline{a}^T\underline{\beta}] + \underline{a}^T(X^TX)^{-1}X^T \xcancel{E[\varepsilon]}^{0} + E[\underline{d}^TX\underline{\beta}] + d^T \xcancel{E[\varepsilon]}^{0}$

$= \underline{a}^T\underline{\beta} + \underline{d}^TX\underline{\beta} = \underline{a}^T\underline{\beta}$ (unbiasedness condition)

$\therefore d^TX = 0$ if and only if unbiased

$Var(\tilde{\theta}) = Var(\underline{c}^T\underline{y}) = \underline{c}^T Var(\underline{y})\underline{c} = \sigma^2\underline{c}^T\underline{c}$

$= \sigma^2(\underline{a}^T(X^TX)^{-1}X^T + \underline{d}^T)(\underline{a}^T(X^TX)^{-1}X^T + \underline{d}^T)^T$

$= \sigma^2(\underline{a}^T(X^TX)^{-1}X^T + \underline{d}^T)(X(X^TX)^{-1}\underline{a} + \underline{d})$

$= \sigma^2(\underline{a}^T\cancel{(X^TX)^{-1}X^TX}(X^TX)^{-1}\underline{a} + \xcancel{d^TX}^{0}(X^TX)^{-1}a$

$+ \underline{a}^T\xcancel{(X^TX)^{-1}X^T\underline{d}}^{0} + d^Td$

$$Var(\tilde{\theta}) = \sigma^2 (\underline{a}^T (X^TX)^{-1} \underline{a} + \underline{d}^T \underline{d})$$

$$Var(\hat{\theta}) = Var(\underline{a}^T (X^TX)^{-1} X^T \underline{y})$$

$$= \underline{a}^T (X^TX)^{-1} X^T \; Var(y) \; (\underline{a}^T (X^TX)^{-1} X^T)^T$$

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

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

$$\therefore \; Var(\tilde{\theta}) = Var(\hat{\theta}) + \underline{d}^T \underline{d}$$

$$\therefore \; Var(\underline{c}^T \underline{y}) = Var(\underline{a}^T \underline{\beta}) + \underline{d}^T \underline{d}$$

$$\therefore \; \underline{Var(\underline{a}^T \underline{\beta}) \le Var(\underline{c}^T \underline{y})}$$

2.

Assume that the column $x_0, \cdots, x_p$ are orthogonal.

$$\hat{\beta} = (X^TX)^{-1} X^T \underline{y} \quad , \quad X = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_p \end{bmatrix}$$

$$= \left(\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_p \end{bmatrix} \begin{bmatrix} x_0 & x_1 & \cdots & x_p \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} x_0 y \\ x_1 y \\ \vdots \\ x_p y \end{bmatrix}$$

$$= \begin{bmatrix} \|x_0\|^2 & x_0 \cdot x_1 & \cdots & x_0 \cdot x_p \\ x_1 \cdot x_0 & \|x_1\|^2 & \cdots & x_1 \cdot x_p \\ \vdots & \vdots & \ddots & \vdots \\ x_p \cdot x_0 & \cdots & \cdots & \|x_p\|^2 \end{bmatrix}^{-1} \begin{bmatrix} x_0 y \\ x_1 y \\ \vdots \\ x_p y \end{bmatrix}$$

$$= \begin{bmatrix} \|x_0\|^2 & 0 & \cdots & 0 \\ 0 & \|x_1\|^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \|x_p\|^2 \end{bmatrix}^{-1} \begin{bmatrix} x_0 y \\ x_1 y \\ \vdots \\ x_p y \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\|x_0\|^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\|x_1\|^2} & \cdots & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \frac{1}{\|x_p\|^2} \end{bmatrix} \begin{bmatrix} x_0 y \\ x_1 y \\ \vdots \\ x_p y \end{bmatrix}$$

$$= \begin{bmatrix} \frac{x_0 y}{\|x_0\|^2} \\ \frac{x_1 y}{\|x_1\|^2} \\ \vdots \\ \frac{x_p y}{\|x_p\|^2} \end{bmatrix}$$

$$\hat{\beta}_J = \sum_{i=0}^{p} \frac{x_i}{\|x_i\|^2} y$$

3.

$x^T x \beta = x^T y$

$X = U \Sigma V^T$

$X \in R_r^{n \times (p+1)}$, $r = \mathrm{rank}(x)$

$U \in R^{n \times r}$ $\quad U^T U = I_r$

$V \in R^{(p+1) \times r}$ $\quad V^T V = I_r$

$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$

(a)

$x^T x \underline{\beta} = x^T \underline{y}$

$$\Rightarrow (U\Sigma V^T)^T (U\Sigma V^T)\underline{\beta} = (U\Sigma V^T)^T \underline{y}$$

$$\Rightarrow V\Sigma^T U^T U \Sigma V^T \underline{\beta} = V\Sigma^T U^T \underline{y}$$

$$\Rightarrow V^T V \Sigma^T U^T U \Sigma V^T \underline{\beta} = V^T V \Sigma^T U^T \underline{y}$$

$$V^T V = I_r \quad U^T U = I_r$$

$$\Rightarrow \Sigma^T \Sigma V^T \underline{\beta} = \Sigma^T U^T \underline{y}$$

$$\Rightarrow (\Sigma^T)^{-1} \Sigma V^T \underline{\beta} = (\Sigma^T)^{-1} \Sigma^T U^T \underline{y}$$

$$\Rightarrow \Sigma V^T \underline{\beta} = U^T \underline{y}$$

$$\Rightarrow \underline{\beta} = \Sigma^{-1} (V^T)^{-1} U^T \underline{y} \qquad (V^T)^{-1} = V$$

$$\underline{\beta} = V\Sigma^{-1} U^T \underline{y}$$

(b)

$$X^T X \underline{\beta} = X^T \underline{y} \qquad \underline{\beta} = V\Sigma^{-1} U^T \underline{y} + b$$

$$(U\Sigma V^T)^T (U\Sigma V^T)(V\Sigma^{-1} U^T \underline{y} + b) = (U\Sigma V^T)^T \underline{y}$$

$$V\Sigma^T U^T U \Sigma V^T V \Sigma^{-1} U^T \underline{y} + V\Sigma^T U^T U \Sigma V^T b = V\Sigma^T U^T \underline{y}$$

$$V\Sigma^T \Sigma \Sigma^{-1} U^T \underline{y} + V\Sigma^T \Sigma V^T b = V\Sigma^T U^T \underline{y}$$

$$V\Sigma^2 V^T b = V\Sigma U^T \underline{y} - V\Sigma^T U^T \underline{y} = 0$$

$\therefore$ if $b=0$, there is only one solution to the normal equation, which is β_{mns}. Otherwise, if $b \neq 0$

$$\|\beta\|^2 = \|\beta_{mns} + b\|^2 = \|\beta_{mns}\|^2 + \|b\|^2$$

$$\Rightarrow \|\beta\| \geq \|\beta_{mns}\|$$

(c)

Prove Penrose properties for $X = U \Sigma V^T$

$X^+ = V \Sigma^{-1} U^T$

1.

$X X^+ X = X$

$$(U \underbrace{\Sigma \overbrace{V^T)(V}^{I} \Sigma^{-1}}_{I} U^T)(U \Sigma V^T) = U \Sigma V^T$$

$$U U^T U \Sigma V^T = U \Sigma V^T$$

$$U \Sigma V^T = U \Sigma V^T$$

proved.

2.

$X^+ X X^+ = X^+$

$$(V \Sigma^{-1} \underbrace{\overbrace{U^T)(U}^{I}}_{I} \Sigma V^T)(V \Sigma^{-1} U^T) = V \Sigma^{-1} U^T$$

$$\underbrace{V V^T}_{I} V \Sigma^{-1} U^T = V \Sigma^{-1} U^T$$

$$V \Sigma^{-1} U^T = V \Sigma^{-1} U^T$$

proved.

3.

$(X X^+)^T = X X^+$

$$(U \Sigma V^T V \Sigma^{-1} U^T)^T = U \Sigma V^T V \Sigma^{-1} U^T$$

$$(U U^T)^T = U U^T$$

$$I = I$$

proved.

4.

$(X^+X)^T = X^+X$

$(V\Sigma^{-1}U^T U\Sigma V^T)^T = (V\Sigma^{-1}U^T U\Sigma V^T)$

$(VV^T)^T = VV^T$

$I = I$ proved.

$\therefore$ $V\Sigma^{-1}U$ is the pseudo inverse of X