CREED REILLY
GESSA HW2
06/08/2021

1) Prove GM Thm

LSE $\rightarrow$ $\hat{\beta} = (X^T X)^{-1} X^T \underline{y}$

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

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

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

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

$$\underline{c}^T = a^T (X^T X)^{-1} X^T + d^T \quad , \underline{y} = X\beta + \underbrace{\underline{\varepsilon}}_{N(0,\sigma^2)}$$

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

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

$$E[\tilde{\theta}] = \underline{a}^T \underline{\beta} + \underline{d}^T X \underline{\beta} = \underline{a}^T \underline{\beta} \quad \text{(unbiased)}$$

$$\therefore (d^T X) = \underline{0} \text{ iff unbiased}$$

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

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

$$= \sigma^2 (a^T (X^T X)^{-1} X^T + d^T)\left[X (X^T X)^{-1} a + d\right]$$

$$= \sigma^2 \left(a^T (X^T X)^{-1} X^T + d^T\right) \left[X (X^T X)^{-1} a + d \right]$$

$$= \sigma^2 \Big[a^T \cancel{(X^T X)^{-1}} \cancel{X^T} X (X^T X)^{-1} a + \underbrace{\cancel{d^T X}}_{\to\, 0 \text{ - unbiased}} (X^T X)^{-1} a$$

$$+ a^T (X^T X)^{-1} \underbrace{X^T d}_{\cancel{(d^T X)^T} \to\, 0 \text{ - unbiased}} + d^T d \Big]$$

$$\text{var}(\tilde{\theta}) = \sigma^2 \left[a^T (X^T X)^{-1} a + d^T d \right]$$

$$\text{var}(\hat{\theta}) = \text{var}\left(a^T (X^T X)^{-1} X y\right)$$

$$= a^T (X^T X)^{-1} X \, \text{var}(y) \left(a^T (X^T X)^{-1} X\right)^T$$

$$= \sigma^2 \left[a^T (X^T X)^{-1} X \right] \left[X^T (X^T X)^{-1} a \right]$$

$$= \sigma^2 \left[a^T (X^T X)^{-1} a \right]$$

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

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

$$\boxed{\therefore \text{var}(\underline{c}^T \underline{y}) \geq \text{var}(\underline{a}^T \underline{\beta})}$$

2) Assume $x_1 \cdots x_p$ are $\perp$ $\leftarrow$ $\text{col}(X) \in \mathbb{R}^n$, full rank

Express $\hat{\beta}_j$ as a $f(x_0 \cdots x_p, \underline{y})$

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

$$\begin{bmatrix} | \\ \hat{\beta} \\ | \end{bmatrix} = \left(\begin{bmatrix} - & x_0 & - \\ - & x_1 & - \\ - & x_2 & - \\ & \vdots & \\ - & x_p & - \end{bmatrix} \begin{bmatrix} | & | & | & & | \\ x_0 & x_1 & x_2 & \cdots & x_p \\ | & | & | & & | \end{bmatrix} \right)^{-1} \begin{bmatrix} | & | & | & & | \\ x_0 & x_1 & x_2 & \cdots & x_p \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} | \\ y \\ | \end{bmatrix}$$

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

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

$$\begin{bmatrix} | \\ \hat{\beta} \\ | \end{bmatrix} = \begin{bmatrix} | & & | \\ \frac{x_0}{\|x_0\|^2} & \cdots & \frac{x_p}{\|x_p\|^2} \\ | & & | \end{bmatrix} \begin{bmatrix} | \\ y \\ | \end{bmatrix}$$

$$\begin{bmatrix} | \\ \hat{\beta} \\ | \end{bmatrix} = \begin{bmatrix} | & & | \\ \frac{x_0}{\|x_0\|^2} & \cdots & \frac{x_p}{\|x_p\|^2} \\ | & & | \end{bmatrix} \begin{bmatrix} | \\ y \\ | \end{bmatrix}$$

$$\boxed{\begin{aligned} \hat{\beta}_j &= \left(\frac{x_0}{\|x_0\|^2}\right)_j y_0 + \left(\frac{x_1}{\|x_1\|^2}\right)_j y_1 + \cdots + \left(\frac{x_p}{\|x_p\|^2}\right)_j y_p \\ \hat{\beta}_j &= \sum_{i=0}^{p} \left(\frac{x_i}{\|x_i\|^2}\right)_j y_i \leftarrow \text{equivalent notation} \end{aligned}}$$

Check:

$$x = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\beta_0 = 1(1) + 0(2) = 1$$

← works out w/ matlab! ✓

$$\beta_1 = 0(1) + \frac{2}{4}(2) = 1$$

3) X^TX not invertible

$X \in \mathbb{R}_r^{n\times(p+1)}$

$r = \text{rank}(X)$

$X = U\Sigma V^T$

$U \in \mathbb{R}^{n\times r}$, $U^TU = I_n$

$\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_r)$

$V \in \mathbb{R}^{(p+1)\times r}$, $V^TV = I_r$

$\Sigma^T = \Sigma$

a. Show that $\beta_{mns} = V\Sigma^{-1}U^T\underline{y}$ is a soln to normal eqns.

Normal Eqns: For $Ax = b$, $A^TAx = A^Tb$

In this case $\rightarrow$ $y = X\beta$

LS soln: $X^TX\beta = X^T\underline{y}$

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

$$(V\Sigma^T\overset{I}{\boxed{U^T)(U}}\Sigma V^T)\beta = V\Sigma^TU^T\underline{y}$$

$$\overset{I}{\boxed{V^T(V}}\Sigma^T\Sigma V^T\underline{\beta}) = \overset{I}{\boxed{V^T(V}}\Sigma^TU^T)\underline{y}$$

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

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

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

$VV^T = I_{p+1}$

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

① $V^TV = I$

$VV^TV = VI = IV$
$= V$

$V[V^TV] = VI$

$[VV^T]V = IV$

$\hookrightarrow VV^T = I_{p+1}$

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

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

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

$$V\Sigma^T \overset{I}{\boxed{\Sigma\Sigma^{-1}}} U^T \underline{y} + V\Sigma^2 V^T b = V\Sigma^T U^T \underline{y}$$

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

only a soln to normal eqn if $b = 0$

If $b \neq 0$, then not a soln to normal eqns, then not minimum!

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

c. Prove Penrose Properties for $X^+ = V\Sigma^{-1}U^T$

$X = U\Sigma V^T$

- Rule 1: $XX^+X = X$

$$LHS = (U\Sigma \boxed{V^T)(V}\Sigma^{-1}U^T)(U\Sigma V^T) = (U\overset{I}{\boxed{\Sigma\Sigma^{-1}}}U^T)(U\Sigma V^T)$$

$$U\overset{I}{\boxed{U^T U}}\Sigma V^T = U\Sigma V^T = RHS \checkmark$$

- Rule 2: $X^+XX^+ = X^+$

$$LHS = (V\Sigma^{-1}\overset{I}{\boxed{U^T)(U}}\Sigma V^T)(V\Sigma^{-1}U^T)$$

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

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

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

- Rule 3: $(AA^+)^T = AA^+$

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

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

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

$$RHS = (U\Sigma V^T)(V\Sigma^{-1}U^T) = UU^T = I \quad \checkmark$$

- Rule 4: $(X^+X)^T = X^+X$

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

$X = U\Sigma V^T$

$LHS = \left(V\Sigma^{-1}U^T U\Sigma V^T\right)^T = V\Sigma^T U^T U(\Sigma^{-1})^T V^T = VV^T = I$

$RHS = (V\Sigma^{-1}U^T U\Sigma V^T) = VV^T = I$ ✓