HW2

Yue Xu

1. Prove the Gauss-Markov Theorem, i.e. show that the least squares estimate in linear regression is the BLUE (Best Linear Unbiased Estimate), which means $\mathrm{Var}(\mathbf{a}^T\widehat{\beta}) \leq \mathrm{Var}(\mathbf{c}^T\mathbf{y})$ where $\mathbf{c}^T\mathbf{y}$ is any unbiased estimator for $\mathbf{a}^T\beta$. (20 pts)

Proof: $\because y = X\beta + \varepsilon$

$\therefore \hat{\beta} = (X^TX)^{-1}X^Ty$
$= (X^TX)^{-1}X^T(X\beta+\varepsilon)$
$= (X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\varepsilon$
$= \beta + (X^TX)^{-1}X^T\varepsilon$

$\therefore E[\hat{\beta}] = E[\beta + (X^TX)^{-1}X^T\varepsilon]$
$= E[\beta] + E[(X^TX^{-1})X^T\varepsilon]$
$= E[\beta] + E[(X^TX^{-1})X^T]\cdot E[\varepsilon]$
$= E[\beta]$ $\quad (\because E[\varepsilon]=0)$

$\therefore \hat{\beta}$ is unbiased —— ①

$\because \mathrm{Var}(a^T\hat{\beta}) = \mathrm{Var}(a^T(X^TX)^{-1}X^T\varepsilon)$
$= a^T E[((X^TX)^{-1}X^T\varepsilon)((X^TX)^{-1}X^T\varepsilon)^T]a$
$= a^T E[(X^TX)^{-1}X^T\varepsilon\varepsilon^T X((X^TX)^{-1})^T]a$
$= a^T(X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1}a$
$= \sigma^2 a^T(X^TX)^{-1}a$ —— ②

$\because c^Ty = c^T(X\beta+\varepsilon)$
$= c^TX\beta + c^T\varepsilon$

$\therefore \mathrm{Var}(c^Ty) = \mathrm{Var}(c^TX\beta + c^T\varepsilon)$
$= \mathrm{Var}(c^T\varepsilon)$
$= c^T\mathrm{Var}(\varepsilon)c$
$= c^T(\sigma^2 I)c$
$= \sigma^2 c^Tc$ —— ③

$\because c^Ty$ is unbiased estimator for $a^T\beta$

$\therefore E[c^Ty] = E[a^T\hat{\beta}]$

$\because E[c^Ty] = c^TE[y] = c^TE[X\beta+\varepsilon] = c^TX\beta$
$E[a^T\hat{\beta}] = a^TE[\hat{\beta}] = a^T\beta$ $\quad$ (from ①)

$\therefore c^TX\beta = a^T\beta$

$\therefore c = (X^T)^{-1}a$

$\therefore \mathrm{Var}(c^Ty) = \sigma^2c^Tc$
$= \sigma^2((X^T)^{-1}a)^T(X^T)^{-1}a$
$= \sigma^2 a^TX^{-1}(X^{-1})^Ta$

$\because a^TX^{-1}(X^{-1})^Ta \geqslant a^T(X^TX)^{-1}a$ $\quad (\because$ Cauchy-Schwarz$)$

$\therefore \mathrm{Var}(c^Ty) \geqslant \mathrm{Var}(a^T\hat{\beta})$ $\quad$ (from ②, ③)

QED

2. (**Linear Regression with Orthogonal Design**) Assume that the columns $\mathbf{x}_0, \ldots, \mathbf{x}_p$ of $\mathbf{X}$ are orthogonal. Express $\widehat{\beta}_j$ in terms of $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_p$ and $\mathbf{y}$. (10 pts)

$\because$ $x_0, \cdots, x_p$ are orthogonal

$$\therefore x_i \cdot x_j = \begin{cases} \|x_i\|^2, & i=j \\ 0, & i \neq j \end{cases}$$

$$\therefore X^T X = \mathrm{diag}(\|x_0\|^2, \|x_1\|^2, \cdots, \|x_p\|^2)$$

$$\therefore (X^T X)^{-1} = \mathrm{diag}\left(\frac{1}{\|x_0\|^2}, \frac{1}{\|x_1\|^2}, \cdots, \frac{1}{\|x_p\|^2}\right)$$

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

$$\therefore \hat{\beta}_j = \frac{x_j^T y}{\|x_j\|^2}, \quad j = 0, 1, \cdots, p$$

3. (**The Minimum Norm Solution**) When $\mathbf{X}^T\mathbf{X}$ is not invertible, the normal equations $\mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y}$ do not have a unique solution. Assume that $\mathbf{X} \in \mathbb{R}_r^{n\times(p+1)}$, where r is the rank of $\mathbf{X}$. Assume that the SVD of $\mathbf{X}$ is $\mathbf{U\Sigma V}^T$, where $\mathbf{U} \in \mathbb{R}^{n\times r}$ satisfies $\mathbf{U}^T\mathbf{U} = \mathbf{I}_r$. Also $\mathbf{V} \in \mathbb{R}^{(p+1)\times r}$ satisfies $\mathbf{V}^T\mathbf{V} = \mathbf{I}_r$ and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_r)$ is the diagonal matrix of positive singular values.

(a) Show that $\beta_{\text{mns}} = \mathbf{V\Sigma}^{-1}\mathbf{U}^T\mathbf{y}$ is a solution to the normal equations. (5 pts)

(b) Show that for any other solution β to the normal equations, $\|\beta\| \geq \|\beta_{\text{mns}}\|$. [Hint: one way (and not the only way) of doing this is to show that $\beta = \beta_{\text{mns}} + b$.] (15 pts)

(c) Is $\mathbf{V\Sigma}^{-1}\mathbf{U}^T$ the pseudo-inverse of $\mathbf{X}$? (Hint: you can prove or disprove using the so-called Penrose properties) (10 pts)

(a) Proof: $X^TX\beta_{mns} = X^TXV\Sigma^{-1}U^Ty$

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

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

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

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

$= (U\Sigma V^T)^Ty$

$= X^Ty$

$\therefore \beta_{mns}$ is a solution to the normal equations

(b) Suppose $\beta = \beta_{mns} + b$, both β and β_{mns} are solutions of equations

$\because X^Ty = X^TX\beta = X^TX\beta_{mns}$

$\therefore X^TX(\beta - \beta_{mns}) = X^TXb = 0$

Let both side $\times\, b^T$

$\therefore b^TX^TXb = (Xb)^TXb = \|Xb\| = 0$

$\therefore Xb = 0$

$\therefore b$ is in the null space of X —— ①

$\because \beta_{mns} = V\Sigma^{-1}U^Ty = V(\Sigma^{-1}U^Ty)$

and first r columns of V are the base of row space of X

$\therefore \beta_{mns}$ is in the row space of X —— ②

$\therefore b$ and β_{mns} are orthogonal (from ①,②)

$\therefore \beta_{mns}^T b = 0$

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

$= (\beta_{mns}+b)^T(\beta_{mns}+b)$

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

$= \|\beta_{mns}\|^2 + \|b\|^2 \geq 0$

$\therefore \|\beta\|^2 \geq \|\beta_{mns}\|^2$

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

(c) Let $X^+ = V\Sigma^{-1}U^T$

i) $XX^+X = U\Sigma V^TV\Sigma^{-1}U^TU\Sigma V^T = U\Sigma I_r\Sigma^{-1}I_r\Sigma V^T = U\Sigma V^T = X$

ii) $X^+XX^+ = V\Sigma^{-1}U^TU\Sigma V^TV\Sigma^{-1}U^T = V\Sigma^{-1}I_r\Sigma I_r\Sigma^{-1}U^T = V\Sigma^{-1}U^T = X^+$

iii) $(XX^+)^T = (U\Sigma V^TV\Sigma^{-1}U^T)^T = U(\Sigma^{-1})^TV^TV\Sigma^TU^T = U\Sigma^{-1}V^TV\Sigma U^T = U\Sigma^{-1}I_r\Sigma U^T = UU^T$

$XX^+ = U\Sigma V^TV\Sigma^{-1}U^T = U\Sigma I_r\Sigma^{-1}U^T = UU^T$

$\therefore (XX^+)^T = XX^+$

iv) $(X^+X)^T = (V\Sigma^{-1}U^TU\Sigma V^T)^T = V\Sigma U^TU\Sigma^{-1}V^T = VV^T$

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

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

According to i) ~ iv), $V\Sigma^{-1}V^T$ is the pseudo-inverse of X