High-dimensional data analysis HW#5

102225014 Chen Ai-Chun

Ex3.3

(a) Prove Gauss-Markov theorem: the least squares estimate of a parameter $a^{\mathrm{T}}\beta$ has variance no bigger than that of any other linear unbiased estimate of $a^{\mathrm{T}}\beta$ (Section 3.2.2).

<u>Solution</u>:

Let $\theta = a^{\mathrm{T}}\beta$ be our target. The least square estimate is $\hat{\theta} = a^{\mathrm{T}}\hat{\beta} = a^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}y$.

And other unbiased estimate is presented as $\tilde{\theta} = a^{\mathrm{T}}\tilde{\beta} = \mathrm{C}_0^{\mathrm{T}}y$.

We have $E(\tilde{\theta}) = \mathrm{C}_0^{\mathrm{T}}X\beta = a^{\mathrm{T}}\beta \Rightarrow \mathrm{C}_0^{\mathrm{T}}X = a^{\mathrm{T}}$.

The variances of $\tilde{\theta}$ and $\hat{\theta}$ are

$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(a^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}y) = \sigma^2 a^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}a$, and

$\mathrm{Var}(\tilde{\theta}) = \mathrm{Var}(C_0^{\mathrm{T}}y) = \sigma^2 C_0^{\mathrm{T}}C_0$.

Therefore, consider

$$\begin{aligned}
\mathrm{Var}(\tilde{\theta}) - \mathrm{Var}(\hat{\theta}) &= \sigma^2 C_0^{\mathrm{T}}C_0 - \sigma^2 a^{\mathrm{T}}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}a \\
&= \sigma^2 C_0^{\mathrm{T}}C_0 - \sigma^2 \mathrm{C}_0^{\mathrm{T}}\mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{C}_0 \\
&= \sigma^2 \mathrm{C}_0^{\mathrm{T}}(I - \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}})\mathrm{C}_0
\end{aligned}$$

Here, let $H = I - \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}$, eigenvalue equation of $H$ is:

$$H\mathbf{x} = \lambda\mathbf{x}.$$

$\Rightarrow H^2\mathbf{x} = \lambda H\mathbf{x}$

$\Rightarrow H\mathbf{x} = \lambda^2\mathbf{x}$

$\Rightarrow \lambda\mathbf{x} = \lambda^2\mathbf{x}$

$\Rightarrow \lambda(1-\lambda)\mathbf{x} = \mathbf{0}$

$\Rightarrow \lambda = 0, 1$

$\Rightarrow H = I - \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}$ is positive semi-definite.

Thus, $\mathrm{Var}(\tilde{\theta}) - \mathrm{Var}(\hat{\theta}) = \sigma^2\mathrm{C}_0^{\mathrm{T}}(I - \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}})\mathrm{C}_0 = \sigma^2\mathrm{C}_0^{\mathrm{T}}\mathrm{H}\mathrm{C}_0 \geq 0$.

(b) The matrix inequality $\mathrm{B} \preceq \mathrm{A}$ holds if $\mathrm{A}-\mathrm{B}$ is positive semi-definite. Show that if $\hat{\mathrm{V}}$ is the variance-covariance matrix of the least squares estimate of $\beta$ and $\tilde{\mathrm{V}}$ is the variance-covariance matrix of any other linear unbiased estimate, then $\hat{\mathrm{V}} \preceq \tilde{\mathrm{V}}$.

Solution:

Let other unbiased estimate of $\beta$ be the form $\tilde{\beta} = Cy = \{ (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}} + D \} y$, where the forward term is the least square and plus $p \times n$ matrix $D$.

Then, by definition of unbiase,

$$
\begin{aligned}
E(\tilde{\beta}) &= \{ (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}} + D \}\mathrm{X}\beta \\
&= (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{X}\beta + D\,\mathrm{X}\beta \\
&= \beta + D\,\mathrm{X}\beta \\
&= \beta
\end{aligned}
$$

We have $\tilde{\beta}$ is unbiased iff $D\mathrm{X} = 0$.

Therefore, consider $\tilde{\mathrm{V}} - \hat{\mathrm{V}}$:

$$
\begin{aligned}
&\tilde{\mathrm{V}} - \hat{\mathrm{V}} \\
&= \mathrm{Var}(\tilde{\beta}) - \mathrm{Var}(\hat{\beta}) \\
&= \sigma^2 CC^{\mathrm{T}} - \sigma^2 (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1} \\
&= \sigma^2 \{ (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}} + D \}\{ (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}} + D \}^{\mathrm{T}} - \sigma^2 (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1} \\
&= \sigma^2 \{ (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1} + (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}D^{\mathrm{T}} + D\mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1} + DD^{\mathrm{T}} \} - \sigma^2 (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1} \\
&= \sigma^2 DD^{\mathrm{T}}
\end{aligned}
$$

Because $\mathbf{x}^{\mathrm{T}}DD^{\mathrm{T}}\mathbf{x} = \left\| \mathbf{x}^{\mathrm{T}}D \right\|^2 \geq 0$ for $\forall \mathbf{x} \in R^p$. Thus, $\tilde{\mathrm{V}} - \hat{\mathrm{V}}$ is positive semi-definite and then $\hat{\mathrm{V}} \preceq \tilde{\mathrm{V}}$.