Wednesday, 28 August 2019 11:04 AM

1st step beyond

VARIABLE SELECTION AND MODEL BUILDING

Given the data how to take a decision.

Assume the true model

$$y = X\beta + \epsilon \quad \text{True model} \qquad \beta \in \mathbb{R}^{k+1},\ y \in \mathbb{R}^n,\ \epsilon \in \mathbb{R}^n$$

$$\epsilon \sim N(0, \sigma^2 I_n)$$

Model is constructed based on p many variables instead of $(k+1)$

$$y = [X_p : X_r]\begin{bmatrix}\beta_p \\ \beta_r\end{bmatrix} + \epsilon \qquad p + r = k+1$$

$$\Rightarrow y = X_p\beta_p + X_r\beta_r + \epsilon$$

$$\beta_r = (\beta_0\ \beta_1 \dots \beta_{k-r}) \qquad \beta_r = (\beta_{k-r+1}, \dots, \beta_{k+1})$$

Fitted $y = X_p\beta_p + \epsilon$

$$\hat\beta_p = (X_p^T X_p)^{-1} X_p^T y \quad \text{Reduced model}$$

$$\hat\beta = (X^T X)^{-1} X^T y \quad \text{Complete model}$$

Q Whether $\hat\beta_p$ is an unbiased estimator or not?

$$\hat\beta_p = (X_p^T X_p)^{-1} X_p y$$

$$E[\hat\beta_p] = E\left((X_p^T X_p)^{-1} X_p^T y\right) = E\left((X_p^T X_p)^{-1} X_p^T [X_p\beta_p + X_r\beta_r]\right)$$

$$= \beta_p + \underbrace{(X_p^T X_p)^{-1}}_{\text{Invertible}}\underbrace{(X_p^T X_r)}_{0}\beta_r$$

This is only possible if every column of X_r is orthogonal to every column of X_p.

Q2 When the estimator $\hat\beta_p$ has less variation compared to whole model?

$$D(\hat\beta_p) = \sigma^2 (X_p^T X_p)^{-1}$$

$$D(\hat\beta) = \sigma^2 (X^T X)^{-1} \quad \text{as } |X^TX|^{-1}$$

$$= \sigma^2\left((X_p : X_r)^T (X_p : X_r)\right)^{-1}$$

$$= \sigma^2\left(\begin{pmatrix}X_p^T \\ X_r^T\end{pmatrix}(X_p\ X_r)\right)^{-1}$$

$$D\begin{pmatrix}\hat\beta_p \\ \hat\beta_r\end{pmatrix} = \sigma^2 \begin{pmatrix}X_p^T X_p & X_p^T X_r \\ X_r^T X_p & X_r^T X_r\end{pmatrix}^{-1} \qquad D(\hat\beta_p)$$

Let us call "A" matrix

$$A^{-1} = \begin{pmatrix}A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix}^{-1}$$

$$= \begin{pmatrix}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -A_{21}A_{11}^{-1}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}\end{pmatrix}$$

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

if $|C^{-1} + DA^{-1}B| \neq 0$

?? Component of $D(\hat\beta)$

$$= \Big[\underbrace{(X_p^T X_p)}_{A} - \underbrace{(X_p^T X_r)}_{B}\underbrace{(X_r^T X_r)^{-1}}_{C}\underbrace{(X_r^T X_p)}_{D}\Big]^{-1}\sigma^2$$

$$= \sigma^2 (X_p^T X_p)^{-1} + \sigma^2\left((X_p^T X_r)^{-1} X_p X_p \left((X_r^T X_r)^{-1} + (X_r^T X_p)(X_p^T X_p)^{-1}(X\right.\right.$$

$$= \sigma^2 (X_p^T X_p)^{-1} + [A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}]\sigma^2$$

?? PSD → Yes

$A = X_p^T X_p \quad B = X_p^T X_r \qquad A^{-1}B = (X_p^T X_p)^{-1} X_p^T X_r = (DA^{-1})^T$

$C = (X_r^T X_r)^{-1} \quad D = X_r^T X_p \qquad DA^{-1} = (X_r^T X_p)(X_p^T X_p)^{-1}$

$$\Rightarrow A^{-1}B(C^{-1} + DA^{-1}B)DA^{-1}$$

$$\hookrightarrow (A^{-1}B)^T$$

$$\Rightarrow A^{-1}B(C^{-1} + DA^{-1}B)(A^{-1}B)^T$$

$$\Rightarrow \underbrace{A^{-1}BC^{-1}B^TA^{-T}}_{PSD} + \underbrace{A^{-1}BDA^{-1}BB^TA^{-T}}_{PSD}$$

If M is a PSD matrix then all eigenvalues are ≥ 0.