ST5227_Tut 3

Zhu Xu

A0191344H

Q1:

$$E\|\hat\beta_{ridge}-\beta\|^2 = E(\hat\beta_{ridge}-\beta)(\hat\beta_{ridge}-\beta)^T = E\hat\beta_{ridge}\hat\beta_r^T - 2E\hat\beta_r\cdot\beta + \beta\beta^T$$

$$\Rightarrow E\hat\beta_r\hat\beta_r^T = E\{(X^TX+\lambda I)^{-1}X^TY\,Y^TX(X^TX+\lambda I)^{-1}\}$$

$$= \frac{1}{(n+\lambda)^2}\cdot E(X^TYY^TX)$$

$$= \frac{1}{(n+\lambda)^2}E\{X^T(X\beta+\varepsilon)(X\beta+\varepsilon)^TX\}$$

$$= \frac{1}{(n+\lambda)^2}E\{X^TX\beta\beta^TX^TX + X^TX\beta\varepsilon^TX + X^T\varepsilon\beta^TX^TX + X^T\varepsilon\varepsilon^TX\}$$

$$= \frac{1}{(n+\lambda)^2}\left[E(n^2\beta\beta^T)+0+0+E(X^T\varepsilon\varepsilon^TX)\right]$$

$$= \frac{n^2\beta\beta^T+p\sigma^2}{(n+\lambda)^2}$$

$$E\hat\beta_r = E\{(X^TX+\lambda I)^{-1}(X^TY)\} = \frac{n}{n+\lambda}\beta$$

$$\Rightarrow E\|\hat\beta_r-\beta\|^2 = \frac{n^2\beta\beta^T+p\sigma^2}{(n+\lambda)^2} - \frac{2n}{n+\lambda}\beta\beta^T+\beta\beta^T = \frac{\lambda^2\beta^T\beta+p\sigma^2}{(n+\lambda)^2}$$

$$\Rightarrow \frac{\partial E\|\hat\beta_r-\beta\|^2}{\partial\lambda} = 0 \Rightarrow \lambda^* = \frac{p\sigma^2}{n\sum_{j=1}^p\beta_j^2}$$

Q2:

Let $Q(\beta_0,\beta_1) = \sum_{i=1}^n w_i(Y_i-\beta_0-\beta_1X_i)^2$

$$\frac{\partial Q(\beta_0,\beta_1)}{\partial\beta_0} = -2\sum_{i=1}^n w_i(Y_i-\beta_0-\beta_1X_i)=0$$

$$\frac{\partial Q(\beta_0,\beta_1)}{\partial\beta_1} = -2\sum_{i=1}^n w_iX_i(Y_i-\beta_0-\beta_1X_i)=0$$

$$\Rightarrow \hat\beta_1 = \frac{(\sum_{i=1}^n w_iY_i)(\sum_{i=1}^n w_iX_i)-\sum_{i=1}^n w_iX_iY_i}{(\sum_{i=1}^n w_iX_i)^2-\sum_{i=1}^n w_iX_i^2},\quad \hat\beta_0 = \sum_{i=1}^n w_iY_i - \hat\beta_1\sum_{i=1}^n w_iX_i$$

Q3:

$$\hat\beta_r = (X^TX+\lambda I)^{-1}X^TY;\quad \hat\beta_L = (X^TX)^{-1}X^TY,\quad X^TX = \sum_{i=1}^n X_iX_i^T = diag(\sum_{i=1}^n X_{i1}^2,\dots,\sum_{i=1}^n X_{ip}^2) = diag(c_1,\dots,c_p)$$

$$\hat\beta_r = diag(\frac{1}{c_1+\lambda},\dots,\frac{1}{c_p+\lambda})X^TY;\quad \hat\beta_L = diag(\frac{1}{c_1},\dots,\frac{1}{c_p})$$

$$\Rightarrow \hat\beta_r = diag(\frac{c_1}{c_1+\lambda},\dots,\frac{c_p}{c_p+\lambda})\cdot\beta_L$$

Q4:

$$Q_1(\beta) = \sum_{i=1}^n(Y_i-\beta_0-\beta_1X_{i1}-\dots-\beta_pX_{ip})^2+\lambda(\beta_1^2+\dots+\beta_p^2)$$

$$Q_2(\beta^c) = \sum_{i=1}^n[Y_i-\beta_0^c-\beta_1^c(X_{i1}-\bar X)-\dots-\beta_p^c(X_{ip}-\bar X)^2] + \lambda\sum_{i=1}^p(\beta_i^c)^2$$

$$\frac{\partial Q_1}{\partial\beta} = \frac{\partial Q_2}{\partial\beta^c} = 0 \Rightarrow \hat\beta_0 = \bar Y-\hat\beta_1\bar X_1-\dots-\hat\beta_p\bar X_p;\quad \hat\beta_0^c = \bar Y$$

centralized data does not influence LSE $\Rightarrow \min Q_1 = \min Q_2$

Q5:

Let $Q(\beta) = \sum_{i=1}^{n} w_i (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_p X_{ip})^2 + \lambda \sum_{i=1}^{p} \beta_i^2$

$$\frac{\partial Q}{\partial \beta_0} = \frac{\partial Q}{\partial \beta_1} = \cdots = \frac{\partial Q}{\partial \beta_p} = 0$$

$$\Rightarrow \begin{cases} 2\sum_{i=1}^{n} w_i (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_p X_{ip}) = 0 \\ \vdots \\ 2\sum_{i=1}^{n} w_i X_{ip} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_p X_{ip}) + 2\lambda\beta_p = 0 \end{cases}$$

$\Rightarrow$ In matrix: $X^T W X \beta + \lambda I_p \beta = X^T W Y$

$\Rightarrow \hat{\beta} = (X^T W X + \lambda I_p)^{-1} X^T W Y$