$y = Bx$

Theoretical part TP supervised learning IMA205

$y = Bx$

(1) $E[\tilde{\beta}] = E[Cy] = E[(H+D)y] = E[Hy] + E[Dy]$

$E[y] = \beta \quad E[y] = \beta H^{-1}$

Assuming an unbiased OLS model, we have that:

$E[Hy] = \beta$

We also have that if $\tilde{\beta}$ is unbiased:

$E[\tilde{\beta}] = \beta = E[Hy] + E[Dy] = \beta + E[Dy] \Rightarrow E[Dy] = 0$

$D\beta H^{-1} = 0$

$Dx = 0$ (2)

Calculating the variance

$Var(Cy) = C(Var(y))C^T = C\sigma^2 C^T = \sigma^2 CC^T$

(Var(y) is a scalar)

$C = H + D$ and $H = (x^Tx)^{-1}x^Ty$, so

$\sigma^2 CC^T = \sigma^2((x^Tx)^{-1}x^T + D)((x^Tx)^{-1}x^T + D)^T$

$\sigma^2 CC^T = \sigma^2(x^Tx)^{-1} + \sigma^2(x^Tx)^{-1}(Dx)^T + \sigma^2(Dx)(x^Tx)^{-1} + \sigma^2 DD^T$

In order for $\tilde{\beta}$ to be unbiased we have that:

$E[Dy] = 0 \Rightarrow E[DX\beta] = 0$

FROM (2)

$Var(\tilde{\beta}) = \sigma^2 CC^T = \sigma^2(x^Tx)^{-1} + \sigma^2 DD^T$

$Var(\tilde{\beta}) = Var(\beta^*) + \sigma^2 DD^T$

$DD^T \geq 0$

② $\beta^*_{ridge} = (x_c^T x_c + \lambda \bar{I})^{-1} x_c^T y_c$

$E[\beta^*_{ridge}] = E[(x_c^T x_c + \lambda \bar{I})^{-1} x_c^T y_c]$

$E[\beta^*_{ridge}] = [(x_c^T x_c + \lambda I)^{-1} x_c^T x_c]\beta$

The SVD decomposition

$\beta^*_{ridge} = (x_c^T x_c + \lambda \bar{I})^{-1} x_c^T y_c = ([UDV^T]^T [UDV^T] + \lambda I)^{-1} (UDV^T)^T y_c$

$\beta^*_{ridge} = V(D^T D + \lambda \bar{I})^{-1} D^T U^T y_c$

The variance of a ridge estimator is given by

$$Var(\beta^*_{ridge}) = \sigma^2 (x^T x_c + \lambda \bar{I})^{-1} x_c^T x_c (x_c^T x_c + \lambda I)^{-1}$$

If $\lambda > 0 \Rightarrow (x_c^T x_c + \lambda I) > x_c^T x_c$

$\Rightarrow (x_c^T x_c + \lambda I)^{-1} < (x_c^T x_c)^{-1}$

$\Rightarrow Var(\beta^*_{OLS}) > Var(\beta^*_{ridge})$

If λ is really close to 0 the model would get similar to a OLS, which means that the bias would be 0

We know that $\beta^*_{ridge} = (x_c^T x_c + \lambda I[d])^{-1} x_c^T y_c$. If $x_c^T x_c = Id$

$\beta^*_{ridge} = (Id + \lambda I[d])^{-1} x_c^T y_c = ((1+\lambda)Id)^{-1} x_c^T y_c$

$\beta^*_{OLS} = \underbrace{(x^T x_c)}_{Id}{}^{-1} x_c^T y_c \Rightarrow \beta^*_{OLS} = x_c^T y_c \Rightarrow \beta^*_{ridge} = \beta^*_{OLS}/(1+\lambda)$

③ $\beta^*_{ElNet} = \arg\min_\beta (y_c - x_c\beta)^T (y_c - x_c\beta) + \lambda_2 \|\beta\|_2^2 + \lambda_1 \|\beta\|_1$

→ Isn't differentiable in 0, but we know that the subgradient for this func in 0 can be defined as

$$\frac{\partial f}{\partial \beta} = \begin{cases} -1, & \beta < 0 \\ 1, & \beta > 0 \\ [-1, 1] & \beta = 0 \end{cases}$$

$x^T x_c = Id \quad \beta^*_{OLS} = x_c^T y_c$

$$2\beta^*_{OLS} - 2\beta(1-\lambda_2) \pm \lambda_1 = 0$$

$$\beta = \frac{\beta^*_{OLS} \pm \frac{\lambda_1}{2}}{(1-\lambda_2)}$$