

Statistical Modelling Assignment 1

Akshat Dwivedi (r0475672)

April 11, 2016

Question 3

Consider the linear model in matrix form $Y = X\beta + \varepsilon$, with $\text{Var}(\varepsilon) = \sigma^2 I_n$, $Y = (Y_1, \dots, Y_n)^t$, $\beta = (\beta_1, \dots, \beta_p)^t$, X an $n \times p$ matrix and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^t$ be the vector of residuals. Consider the ridge regression estimator $\hat{\beta}_R = (X^t X + \lambda I_p)^{-1} X^t Y$.

Part (a)

Compute the bias of the estimator $\hat{\beta}_R$ using that for $A = (I_p + \lambda(X^t X)^{-1})^{-1}$, it holds that $\hat{\beta}_R = A\hat{\beta}_{LS}$, with $\hat{\beta}_{LS}$ the least squares estimator of β .

Proof. Consider the ridge regression estimator

$$\hat{\beta}_R = (X^t X + \lambda I_p)^{-1} X^t Y \quad (0.1)$$

Using $I = (X^t X)(X^t X)^{-1}$ and substituting into the equation above gives us

$$\hat{\beta}_R = (X^t X + \lambda(X^t X)(X^t X)^{-1})^{-1} X^t Y \quad (0.2)$$

Taking $(X^t X)$ common and using $(AB)^{-1} = B^{-1}A^{-1}$ we get

$$\hat{\beta}_R = (I_p + \lambda(X^t X)^{-1})^{-1} (X^t X)^{-1} X^t Y \quad (0.3)$$

Substituting $Y = X\beta + \varepsilon$ into the previous equation gives us

$$\hat{\beta}_R = (I_p + \lambda(X^t X)^{-1})^{-1} (X^t X)^{-1} X^t (X\beta + \varepsilon) \quad (0.4)$$

$$= (I_p + \lambda(X^t X)^{-1})^{-1} \beta + (I_p + \lambda(X^t X)^{-1})^{-1} (X^t X)^{-1} X^t \varepsilon \quad (0.5)$$

because $(X^t X)^{-1} (X^t X) = I$. Taking the expectation, we get

$$E(\hat{\beta}_R) = E[(I_p + \lambda(X^t X)^{-1})^{-1} \beta] \quad (0.6)$$

$$= (I_p + \lambda(X^t X)^{-1})^{-1} E(\beta) \quad (0.7)$$

$$= (I_p + \lambda(X^t X)^{-1})^{-1} \beta \quad (0.8)$$

since $E(\varepsilon) = 0$ and using the least squares property that $E(\beta_{LS}) = \beta_{LS}$ since least squares is an unbiased estimator of β . Using $\text{bias}(\hat{\beta}_R) = E(\hat{\beta}_R) - \beta$, we get

$$\text{bias}(\hat{\beta}_R) = ((I_p + \lambda(X^t X)^{-1})^{-1} - I_p) \beta \quad (0.9)$$

which is the desired expression for $\text{bias}(\hat{\beta}_R)$. \square

Part (b)

Compute the variance of the estimator $\hat{\beta}_R$. Note that variance is a matrix.

Proof. We use the result $\hat{\beta}_R = A\beta_{LS}$ where $A = (I_p + \lambda(X^t X)^{-1})^{-1}$ from part(a) above. The expression for variance is derived as follows:

$$\text{Var}(\hat{\beta}_R) = \text{Var}(A\beta_{LS}) \quad (0.10)$$

$$= A\text{Var}(\beta_{LS})A^t \quad (0.11)$$

$$= \sigma^2 A(X^t X)^{-1} A^t \quad (0.12)$$

using the fact that $\text{Cov}(AX, BY) = A\text{Cov}(X, Y)B^t$ and substituting the value of A which results in the desired expression

$$\text{Var}(\hat{\beta}_R) = \sigma^2 (I_p + \lambda(X^t X)^{-1})^{-1} (X^t X)^{-1} ((I_p + \lambda(X^t X)^{-1})^{-1})^t \quad (0.13)$$

\square

Part (c)

Finally, we derive the expressions for the bias and variance of \hat{Y}_R , where $\hat{Y}_R = X\hat{\beta}_R$.

Proof. The expression for bias is obtained as follows:

$$E[\hat{Y}_R] = E[X\hat{\beta}_R] \quad (0.14)$$

$$= X E[\hat{\beta}_R] \quad (0.15)$$

$$= X(I_p + \lambda(X^t X)^{-1})^{-1} \beta \quad (0.16)$$

using $\text{bias}(\hat{Y}_R) = E[\hat{Y}_R] - E[Y]$

$$\text{bias}(\hat{Y}_R) = X(I_p + \lambda(X^t X)^{-1})^{-1}\beta - X\beta \quad (0.17)$$

$$= X((I_p + \lambda(X^t X)^{-1})^{-1} - I_p)\beta \quad (0.18)$$

The expression for $\text{Var}(\hat{Y}_R)$ is derived as follows

$$\text{Var}(\hat{Y}_R) = \text{Var}(X\hat{\beta}_R) \quad (0.19)$$

$$= \text{Var}(XA\hat{\beta}_{LS}) \quad (0.20)$$

$$= (XA)\text{Var}(\hat{\beta}_{LS})(XA)^t \quad (0.21)$$

$$= \sigma^2(XA)(X^t X)(XA)^t \quad (0.22)$$

where $A = (I_p + \lambda(X^t X)^{-1})^{-1}$.

□