Statistical Modelling Assignment 1

Akshat Dwivedi ([redacted])

April 11, 2016

Question 3

Consider the linear model in matrix form $Y = X\beta + \varepsilon$, with $\texttt{Var}(\varepsilon) = \sigma^2 I_n$, $Y = (Y_1, \ldots, Y_n)^t$, $\beta = (\beta_1, \ldots, \beta_p)^t$, X an n x p matrix and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^t$ 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

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

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

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

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

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

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

$$\begin{aligned} \hat{\beta}_R &= (I_p + \lambda(X^t X)^{-1})^{-1}(X^t X)^{-1} X^t (X\beta + \boldsymbol{\varepsilon}) && (0.4) \\ &= (I_p + \lambda(X^t X)^{-1})^{-1}\beta + (I_p + \lambda(X^t X)^{-1})^{-1}(X^t X)^{-1} X^t \boldsymbol{\varepsilon} && (0.5) \end{aligned}$$

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

$$E(\hat{\beta}_R) = E[(I_p + \lambda(X^tX)^{-1})^{-1}\beta] \tag{0.6}$$
$$= (I_p + \lambda(X^tX)^{-1})^{-1}E(\beta) \tag{0.7}$$
$$= (I_p + \lambda(X^tX)^{-1})^{-1}\beta \tag{0.8}$$

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

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

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

Part (b)

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

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

$$Var(\hat{\beta}_R) = Var(A\hat{\beta}_{LS}) \tag{0.10}$$
$$= AVar(\hat{\beta}_{LS})A^t \tag{0.11}$$
$$= \sigma^2 A(X^tX)^{-1}A^t \tag{0.12}$$

using the fact that $Cov(AX, BY) = ACov(X, Y)B^t$, $\text{Var}(\hat{\beta}_{LS}) = \sigma^2(X^tX)^{-1}$, and substituting the value of A which results in the desired expression

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

□

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:

$$E[\hat{Y}_R] = E[X\hat{\beta}_R] \tag{0.14}$$
$$= XE[\hat{\beta}_R] \tag{0.15}$$
$$= X(I_p + \lambda(X^tX)^{-1})^{-1}\beta \tag{0.16}$$

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

$$\begin{aligned}
\text{bias}(\hat{Y}_R) &= X(I_p + \lambda(X^tX)^{-1})^{-1}\beta - X\beta && (0.17)\\
&= X((I_p + \lambda(X^tX)^{-1})^{-1} - I_p)\beta && (0.18)
\end{aligned}$$

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

$$\begin{aligned}
\text{Var}(\hat{Y}_R) &= \text{Var}(X\hat{\beta}_R) && (0.19)\\
&= \text{Var}(XA\hat{\beta}_{LS}) && (0.20)\\
&= (XA)\text{Var}(\hat{\beta}_{LS})(XA)^t && (0.21)\\
&= \sigma^2(XA)(X^tX)^{-1}(XA)^t && (0.22)
\end{aligned}$$

using $\text{Var}(\hat{\beta}_{LS}) = \sigma^2(X^tX)^{-1}$ and where $\text{A} = (I_p + \lambda(X^tX)^{-1})^{-1}$.

$\square$