Homework 5 (Due on November 27th midnight)

Total marks 40

From Chapter 6 page 261 (Use R or Rstudio)

Q5. It is well-known that ridge regression tends to give similar coefficient values to correlated variables, whereas the lasso may give quite different coefficient values to correlated variables. We will now explore this property in a very simple setting. Suppose that $n = 2, p = 2, x_{11} = x_{12}, x_{21} = x_{22}$. Furthermore, suppose that $y_1 + y_2 = 0$ and $x_{11} + x_{21} = 0$ and $x_{12} + x_{22} = 0$, so that the estimate for the intercept in a least squares, ridge regression, or lasso model is zero: $\hat{\beta}_0 = 0$.

(a) **(3 marks)** Show that the ridge regression optimization problem in this setting (or the quantity in equation 6.5 in Chapter 6 in this setting) is $2(y_1 - (\beta_1 + \beta_2)x_{11})^2 + \lambda({\beta_1}^2 + {\beta_2}^2)$.

(b) **(5 marks)** Show that in the setting (a), the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$.

(c) **(3 marks)** Show that the lasso regression optimization problem in this setting (or the quantity in equation 6.7 in Chapter 6 in this setting) is $2(y_1 - (\beta_1 + \beta_2)x_{11})^2 + \lambda(|\beta_1| + |\beta_2|)$.

(d) **(5 marks)** Show that in the setting (c), the lasso coefficients $\hat{\beta}_1$and $\hat{\beta}_1$are not unique—in other words, there are many possible solutions to the optimization problem.

Q8. In this exercise, we will generate simulated data, and will then use this data to perform best model selection. Use the $rnorm()$ function to generate a predictor X of length $n = 100$, as well as a noise vector ϵ of length n = 100 such that ϵ = 0.1 * $rnorm(n)$

(a) **(1 mark)** Generate **(use set.seed(19))** a response vector Y of length $n = 100$ according to the model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$$

where β_0, β_1, β_2, and β_3 are constants as $\beta_0 = 1.0$, $\beta_1 = -0.1$, $\beta_2 = 0.05$, $\beta_3 = 0.75$

(b) Use the regsubsets() function to perform **best subset selection** in order to choose the best model containing the predictors $X, X^2, X^3, \ldots, X^8$ using the measures $\boldsymbol{C_p, BIC, adjusted\ R^2}$

(i) **(6 marks)** Plot each measure against number of predictors on the same page using par(mfrow=c(2,2)).
(ii) **(3 marks)** Give the best model coefficients obtained from each $C_p, BIC, adjusted\ R^2$.

Note:

1. You will need to use the data.frame() function to create a single data set containing both X and Y.

(c) Now fit a **ridge regression model** to the simulated data, again using $X, X^2, X^3, \ldots, X^8$ as predictors.

(i) **(2 marks)** Plot the extracted coefficients as a function of log(λ) with a legend containing each curve colour and its predictor name at the top-right corner.

(ii) **(4 marks)** Plot the cross-validation (**set.seed(20))** error as a function of log(λ) to find the **optimal λ .**

(iii) **(1 mark)** Give coefficient estimates for the optimal value of λ.

(d) Now fit a **lasso model** to the simulated data, again using $X, X^2, X^3, \ldots, X^8$ as predictors.

(i) **(2 marks)** Plot the extracted coefficients as a function of log(λ) with a legend containing each curve colour and its predictor name at the top-right corner.

(ii) **(4 marks)** Plot the cross-validation (**set.seed(21))** error as a function of log(λ) to find the **optimal λ .**

(iii) **(1 mark)** Give coefficient estimates for the optimal value of λ.

Note:

1. Use cv.glmnet() to do the cross-validation and use the default of 10-fold cross-validation.