Statistical Learning and Analysis
Instructor: Prof. Varun Rai
Assignment 4

Question 1

Read and summarize the subsection on "Another Formulation for Ridge Regression and the Lasso" (pp 220-222).

Question 2

In this exercise, we will generate simulated data, then use this data to perform forward feature selection, backward feature selection and lasso.

(a) Generate a predictor X of length n = 100, as well as a noise vector ϵ of length n = 100 from a random normal distribution. Then generate 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 β0, β1, β2, and β3 are constants of your choice. Hint: Look back at Assignment 1.

(b) Using forward stepwise selection and also using backwards stepwise selection to choose the best model containing the predictors X, X^2, ... X^6. Comment on your results.

(c) Now fit a lasso model to the simulated data, again using X, X^2, ... X^6 as predictors. Use cross-validation to select the optimal value of λ. Create plots of the cross-validation error as a function of λ. Report the resulting coefficient estimates and discuss the results obtained.