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

Question 1:
Read and summary the subsection on "Another Formulation for Ridge Regression and the Lasso" (pp 220-222)

Question 2:

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

(a) Generate a predictor X of length n = 100, as well as a noise vector ϵ of length n = 100 from 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 $\beta 0$, $\beta 1$, $\beta 2$, and $\beta 3$ are constants of your choice.

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

Question 3:

Draw an example (of your own invention) of a partition of two-dimensional feature space that could result from recursive binary splitting. Your example should contain at least six regions. Draw a decision tree corresponding to this partition. Be sure to label all aspects of your figures, including the regions R1, R2, . . . , the cut points t1, t2, . . . , and so forth.