**DSI -06 Homework 5:**

Introduction with Statistical Learning with Applications to Python (ISLP) pg. 286-287

9. In this exercise, we will predict the number of applications received using the other variables in the College data set.

1. Split the data set into a training set and a test set.
2. Fit a linear model using least squares on the training set, and report the test error obtained.
3. Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.
4. Fit a lasso model on the training set, with λ chosen by cross validation. Report the test error obtained, along with the number of non-zero coefficient estimates.
5. Fit a PCR model on the training set, with *M* chosen by cross validation. Report the test error obtained, along with the value of M selected by cross-validation.
6. Fit a PLS model on the training set, with *M* chosen by cross validation. Report the test error obtained, along with the value of M selected by cross-validation.
7. Comment on the results obtained. How accurately can we predict the number of college applications received? Is there much difference among the test errors resulting from these five approaches?