Assignment #5: Advanced Regression

Problem 1

In this question, we will predict the number of applications received (Apps) using the other variables in the College data set (ISLR package).

(a) Perform best subset selection to the data. What is the best model obtained according to C_p, BIC and adjusted R^2? Show some plots to provide evidence for your answer, and report the coefficients of the best model.

(b) Repeat (a) using forward stepwise selection and backwards stepwise selection. How does your answer compare to the results in (a)?

(c) Fit a lasso model on the data. 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.

(d) Fit a ridge regression model on the data. 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.

(e) Now split the data set into a training set and a test set.

i. Fit the best models obtained in the best subset selection (according to C_p, BIC or adjusted R^2) to the training set, and report the test error obtained.

ii. Fit a lasso model to the training set, with λ chosen by cross validation. Report the test error obtained.

iii. Fit a ridge regression model to the training set, with λ chosen by cross validation. Report the test error obtained.

iv. Compare the test errors obtained in the above analysis (i-iii) and determine the optimal model.

Submit through link: eCampus -> Assignments-> Assignment 5 submission

Deadline: Oct 31, Tue @11:59pm