

# HW8

Math 189

Friday of Week 8, 05/20/2022

1. We will predict the number of applications received (**Apps**) using the other variables in the **College** data set (use the **ISLR2** library).
  - (a) Randomly split the data set into a training set and a test set. You may choose your own random seed.
  - (b) Fit a linear model using least squares on the training set, and report the test error obtained.
  - (c) Fit a ridge regression model on the training set, with  $\lambda$  chosen by cross-validation. Report the test error obtained.
  - (d) Fit a lasso model on the training set, with  $\lambda$  chosen by cross-validation. Report the test error obtained, along with the number of non-zero coefficient estimates.
  - (e) 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.
2. This question uses the variables **dis** (the weighted mean of distances to five Boston employment centers) and **nox** (nitrogen oxides concentration in parts per 10 million) from the **Boston** data. We will treat **dis** as the predictor and **nox** as the response.
  - (a) Use the **poly()** function to fit a cubic polynomial regression to predict **nox** using **dis**. Report the regression output, and plot the resulting data and polynomial fits.
  - (b) Plot the polynomial fits for a range of different polynomial degrees (say, from 1 to 10), and report the associated residual sum of squares.
  - (c) Perform cross-validation or another approach to select the optimal degree for the polynomial, and explain your results.