Homework 2 part b

Question 1

Do #4 in section 6.8 of the Introduction to Statistical Learning book.

Question 2

Do #1 in section 7.9 of the book.

Question 3

This question involves the dataset files “hw2simdatatrain” and “hw2simdatatest”. The train file contains “traindata”, a matrix with the variables on the columns, and “y”, a vector of responses corresponding to the rows of traindata. The test file only contains “testdata”, a new matrix of variables like before, but no “y” variable. This is your task:

1. Train the best possible prediction model using “traindata” and “y”. You should consider all of the three following regression techniques: OLS, ridge regression, and lasso regression.
2. Report the model you decided to use. Describe the process you followed for deciding which to use. This description should be thorough.
3. Using “testdata”, produce a vector called “ypred” and save this in a “.Rdata” file with the title “ypred_simdata.Rdata”. Your predictions will be compared against the true values.

Question 4

For this question, you will use the data in the file “oilimportstrain.Rdata”. which contains the matrix “train” with rows “CPI” (Consumer price index) and “TotVal1000” (Total value of oil imports, in $1000s), and the file “oilimportstestCPI.Rdata”, which contains a vector called “testCPI”. This is your task:

1. Train the best possible model using one of the techniques from chapter 7 of Intro to Statistical Learning.
2. Report the model you decided to use. Describe the process you followed for deciding which to use. This description should be thorough.
3. Using “testCPI”, produce a vector called “ypred” and save this in a “.Rdata” file with the title “ypred_oil.Rdata”. Your predictions will be compared against the true values.