

## STATS 415 Homework 7

Due Thursday March 15, 2018

**Please include your name, username, and lab section (number or time or GSI). A point will be taken off homework without the section info.** Turn in a printout of your homework in the lecture or in your GSI's mailbox across room 305A West Hall, no later than 5pm on the due date.

This homework continues Homework 6, with the same task of predicting the number of applications received using the other variables in the `College` data set. Use the exact same split into training and test data as you used in Homework 6.

1. Perform Principal Component Analysis on the predictors. Make a scree plot of the eigenvalues. How many eigenvalues does one need to explain 90% of the variance in the data? **Report loadings of the first two PCs. Interpret them if you can.**
2. Fit a PCR model on the training set, with the number of principal components  $K$  chosen by cross-validation. Report the training and test error obtained, along with the value of  $K$  selected.
3. Fit a PLS model on the training set, with the number of principal components  $K$  chosen by cross-validation. Report the training and test error obtained, along with the value of  $K$  selected.
4. Comment on the results obtained, including also the methods from homework 6. Which approach would you recommend for this dataset and why?

Please limit your solution to at most 4 pages.