BSAD 646: Statistical Learning

Midterm 1: Exploratory Data Analysis, Regression, Classification

Spring 2018

For each problem, show your code and your results.

1. A regression problem

On the Google Drive, in the Data folder, there is a subfolder called Midterm 1. In it, there are five CSV files about the NBA draft. In those files, the column named "Draft pick" is empty if the player was not drafted. If there is a number, it indicates their place in the draft. (For instance, lower numbers were drafted first.) Use these files–*and only the players that were drafted* to develop a multiple regression model predicting place in the draft based on other characteristics. [1]

2. A classification problem

Now using *all* the data, develop a model that predicts whether someone was drafted or not. (E.g., ignore their position in the draft, we just want to predict Drafted = 1 or Drafted = 0.)[2]

This is due Friday at 5:00 PM, on Blackboard.

Don't forget, you get one free consult with me during office hours, classtime, or over email. (General questions about the data don't count, needing help reading in the files would count.)

[1] Hint: show some exploratory analysis through plots, correlations, and so on. Don't forget to diagnose potential problems. Make any reasonable assumptions about the data you feel you need. For both problems, don't just show your final model, show the steps you took to reach it. Recall, a good model probably doesn't include every predictor, particularly if the predictor is not significant!

[2] Be sure to split the data into a training and test dataset. Develop your model using the training data, and then show how your chosen model performs on the testing data.