BUDA 525 Final— Group Assignment– DUE OCT 8, 2019 by 11:59PM

Brad Price

9/23/2019

## Directions

This is the final group project for BUDA 525. Submit both the .Rmd and word document to the appropriate link on ecampus by 11:59 on Tuesday October 8, 2019. There are 4 questions on this assignment. The first two questions are discussion type questions which require no data analysis but do require substantial insights into the material that has been covered. The final two questions are data analysis type questions which require you to analyze data and communicate relevant findings. You should only show results in this document that relevant to your discussions. Remember you can use ECHO=FALSE, and other R markdown arguments to present your code. Since I have the .Rmd I can see all code that was not presented in the word document.

## Problem 1 (20 points)

Compare and contrast the bootstrap and randomization test procedures we have discussed. In this discussion you should describe the purpose of each method (why would you use them), the procedure they follow, and then how they differ. Finally give a few sentences on each that would explain to a novice what these methods do and why we should trust their results.

## Problem 2 (20 Points)

Compare and contrast the model selection procedures we’ve discussed in this course. Specifically focus your discussion on R-squared, Anova, AIC, BIC, K-Fold Cross Validation, and Random Splitting. You should further focus on when are these methods able to be used and when are they not. Finally discuss your preference for a technique and give reasons why.

## Problem 3 (30 Points)

In the Credit data in the ISLR package it contains 10,000 customers and information on their credit history. For full information of the data look at the help file. A company has approached us to better understand factors that influence the Balance variable which is average credit card balance in USD. Using the information in the model discuss the influential factors, and discuss the factors you choose to put in the model. Do you have any concerns about the use of certain variables in the model? Discuss how your model was created and any insights you can provide based on the results. HINT: Adding Gender and/or Ethnicity could be controversial or illegal in some uses of this this model you should discuss your decision on these variables and how it effects the organizations ability to use your model for prediction or inference.

library(ISLR)  
head(Credit)

## ID Income Limit Rating Cards Age Education Gender Student Married  
## 1 1 14.891 3606 283 2 34 11 Male No Yes  
## 2 2 106.025 6645 483 3 82 15 Female Yes Yes  
## 3 3 104.593 7075 514 4 71 11 Male No No  
## 4 4 148.924 9504 681 3 36 11 Female No No  
## 5 5 55.882 4897 357 2 68 16 Male No Yes  
## 6 6 80.180 8047 569 4 77 10 Male No No  
## Ethnicity Balance  
## 1 Caucasian 333  
## 2 Asian 903  
## 3 Asian 580  
## 4 Asian 964  
## 5 Caucasian 331  
## 6 Caucasian 1151

## Problem 4 (30 Points)

The Salaries data in the carData package contains information on academic salaries in 2008 and 2009 in a college in the US. A data dictionary can be found in the help file for the data. This data was collected as part of an on-going effort of the college to monitor salary differences between male and female faculty members. We have been asked to investigate the gender gap in the data, but also what other information that may be relevant to admistrators (i.e. salary growth for years of service, discipline based growth, etc). Investigate if there is a gender gap, but also provide insights on other drivers that you may see of salary in the data. Is your model suitable to make offers based on the infromation provided? Explain your reasoning. Provide insights into any other information you find of interest.

library(carData)  
head(Salaries)

## rank discipline yrs.since.phd yrs.service sex salary  
## 1 Prof B 19 18 Male 139750  
## 2 Prof B 20 16 Male 173200  
## 3 AsstProf B 4 3 Male 79750  
## 4 Prof B 45 39 Male 115000  
## 5 Prof B 40 41 Male 141500  
## 6 AssocProf B 6 6 Male 97000