Yeabin Moon, Ph.D.
BUS 212A-2 Spring 2023
January 5, 2023

Homework 5

Submission Instructions

1) You have to use Juppyter Notebook
2) Click the Save button at the top of the Jupyter Notebook.
3) Select Cell → All Output → Clear. This will clear all the outputs from all cells (but will keep the content of all cells).
4) Select Cell → Run All. This will run all the cells in order, and will take several minutes.
5) Once you've rerun everything, select File → Download as → PDF via LaTeX (If you have trouble using "PDF via LaTex", you can also save the webpage as pdf. Make sure all your solutions especially the coding parts are displayed in the pdf, it's okay if the provided codes get cut off because lines are not wrapped in code cells).
6) Look at the PDF file and make sure all your solutions are there, displayed correctly. The PDF is the only thing your graders will see!
7) Submit your PDF on Latte.

Question 1. If a decision tree is overfitting the training set, is it a good idea to try decreasing **max_depth** parameter?

Question 2. If a decision tree is underfitting the training set, is it a good idea to try scaling the input features?

Question 3. How can decision tree algorithms be used to optimize inventory management in a retail business?

Question 4. In class, we applied random forests to the Boston data using $mtry = 6$ and using $ntree = 25$ and $ntree = 500$. Create a plot displaying the test error resulting from random forests on this data set for a more comprehensive range of values for $mtry$ and $ntree$. Describe the results obtained.

Question 5. Apply boosting, bagging, random forests, and BART to a data set of your choice. Be sure to fit the models on a training set and to evaluate their performance on a test set. How accurate are the results compared to simple methods like linear or logistic regression? Which of these approaches yields the best performance?