DATA 5610: Homework 1

February 28, 2022

Introduction

This exercise is to help you learn how to implement and execute your first classification algorithm.

Problem 1.

1. Using `pandas` load the `islr2_default` data into a `DataFrame`

2. Train a Bayesian classifier using the `income` variable.

3. Split the original dataset into two subsets:

    - Training dataset: randomly select 80% for Training
    - Test dataset: the remaining 20% for testing

4. Once you have trained the model, make predictions on the observations in the test dataset.

5. Calculate the error rate versus the *"known"* outcome (these are simulated data)

6. Make a table to present your results

Problem 2.

1. Repeat the above exercise for the `islr2_smarket.csv` dataset

2. Use the `volume` variable as your feature/explanatory/predictor variable

3. Make a table presenting your results

4. What assumptions are we making in this example about the data that probably do not hold? How would you revise the model to make more appropriate for these data?