Ec 5043 Prof. Le Wang
Managerial Economics II

Homework #3

February 6, 2019

Instruction: Do all the following empirical exercises using R. Turn in your answer with tables and graphs, if any, (along with your program and output files appended at the end of document). Refer to the output file whenever appropriate when discussing your results.

Note that for all simulation exercises, set the seed number to 123456 to ensure the reproducibility of your results.

1. Question 1. [Classification and Bayes Classifier]

Use the **Default** data in the ISLR package used in class.

1. Discretize the **balance** variable into four groups and called it **balance.group**: 1 if balance is less than 481.7, 2 if balance is betwen 481.7 and 823.6, 3 if balance is between 823.6 and 1166.3, and 4 if balance is greater than 1166.3.
2. Calculate the following conditional distribution

$$\Pr[\text{default} \mid \text{student} = \text{Yes}, \text{balance} = 3]$$

   (a) Use the subset approach
   (b) Obtain conditional distribution using the joint distribution of all three variables and the marginal distribution of the predictors.

3. What is your classification or prediction for default when someone is a student with the balance level of 3?
4. Let's use **naive_bayes()** command and **predict()** command to generate a prediction. Note that here we have more than two predictors, the Naive Bayes classifier imposes additional assumptions. But it is okay. I just want to see if you know how to run the command and create a new dataset with information that someone is a student with the balance level of 3.