EPPS6323 Knowledge Mining

Assignment 6

1. Lab08 in R
2. Review ISLR Chapters 5 and look up answers for the following questions
   a. What is the difference between Cross-validation and Bootstrap?
   b. How about Jackknife and Bootstrap?
   c. What is the precautionary step of CV?
3. Compare CV and Bootstrap (by sample size).
4. Application exercise:

Use the Default data set to fit a logistic regression model that uses income and balance to predict default.

1. Using the validation set approach, estimate the test error of this model. In order to do this, perform the following steps:
   i. Split the sample set into a training set and a validation set.
   ii. Fit a multiple logistic regression model using only the training observations.
   iii. Obtain a prediction of default status for each individual in the validation set by computing the posterior probability of default for that individual, and classifying the individual to the default category if the posterior probability is greater than 0.5.
   iv. Compute the validation set error, which is the fraction of the observations in the validation set that are misclassified.