

PGDM/PGDM(IB), 2017-19
Predictive Business Analytics
Subject Code- DM-341/IB-316

Time: 2 hrs. 30 mins.

Max Marks: 50 Trimester-III, End-Tem Examination: March 2018

For this examination you have to work on your laptop, using R. Your answers, all the code that you use, as well as the results will have to be written on the Answer Script provided.

Marks will be awarded for clarity and completeness in answers.

Instructions

The dataset to be used for this part is 'german_credit.csv'. The data provides Creditability classification; 1 for Credit-worthy and 0 otherwise. All questions are based on this dataset.

1 Decision Tree

- 1.1 Write a routine to split the dataset into training and test in the ratio 50:50. Ensure that each dataset contains a representative proportion of Creditability records - i.e. proportion close to that in the full dataset.
- 1.2 Create a decision tree model using *rpart*. What is the prediction accuracy using this decision tree model for the training set, and for the test set? Write down the *Confusion matrix* for the test set.
- 1.3 Prune the tree using an appropriate cost-complexity parameter. What is the prediction accuracy of the pruned tree on the test data? Write down the new Confusion Matrix based on this pruned tree.
- 1.4 Repeat the previous part 10 times for ten 50:50 partitions of the full dataset. Note down the prediction accuracies on the test data in each iteration. What is the mean and the standard deviation of the accuracies?

2 k-NN Classification

- 2.1 Split the dataset into training and test in 60:40 ratio. Ensure that each dataset contains a representative proportion of Creditability records - i.e. proportion close to that in the full dataset.
- 2.2 Perform a k-NN Classification for k varying from 10 to 20 in steps of 2. Note down the accuracy in each case.
- 2.3 Repeat k-NN Classification 10 times for the best k from question [2.2]. Note down the prediction accuracies on the test data in each iteration. What is the mean and the standard deviation of the accuracies?

66