Activity Sheet

Learning outcomes

After solving these exercises, you should be able to understand the following:

1. Applying the CART ,C5.0 Decision Tree algorithms to solve classification and regression problems

2. Interpreting the results generated from each algorithm in R

3. Comparison of the regression model performance in terms of MSE and RMSE , etc

4. Comparison of the classification model performance in terms of Precision, Recall and Accuracy

Dataset:

The Univ_bank dataset has 12 variables and 5000 records. Use loan as target variable for classification problem and income as target variable for regression problem.

Steps to follow:

1. Import the data into R

2. Drop the attributes which are not required for model building and convert the nominal attributes into factors

3. Split the data into train, test and evaluation data sets

4. Install and load 'rpart' library

5. Build regression model using method 'anova' on train data

6. Predict the values using fitted model on train, test and evolution data sets

7. Interpret the results (Decision tree plot, pruning the tree, error metrics)

8. Build classification model using method 'class' on train data

9. Predict the values using fitted model on train, test and evolution data sets

10. Interpret the results, construct classification matrix and compare error metrics

2.

1. Import the data into R

2. Drop the attributes which are not required for model building and convert the nominal attributes into factors

3. Split the data into train, test and evaluation data sets

4. Install and load 'C50' library

5. Build classification model using C50 library

6. Interpret the results (exploring the rules, importance of attributes)

7. Predict the values using fitted model on train, test and evolution data sets

8. Construct classification matrix and compare error metrics