RMIT University
School of Science
COSC2110/COSC2111 Data Mining
Laboratory Week 11

Aims of this lab

- Investigate the benefits of ensemble classifiers generated by bagging, boosting and RandomForest.

This lab requires the file:
`data/arff/credit-g.arff`
in the directory
`/KDrive/SEH/SCSIT/Students/Courses/COSC2111/DataMining`

1. Load the file into weka. Using J48 get the classification accuracy on the base file.
2. From the meta classifiers select Bagging. Run the Bagging classifier for different numbers of iterations, for example 5,10,100,200,500,1000,2000,4000 and build a table of results.

   What do you observe?
3. From the meta classifiers select AdaBoostM1. Run the AdaBoostM1 classifier for different numbers of iterations, for example 5,10,100,200,500,1000,2000,4000 and build a table of results.

   What do you observe?
4. From the tree classifiers select RandomForest. Run the RandomForest classifier for different numbers of iterations, for example 5,10,100,200,500,1000,2000,4000 and build a table of results.

   What do you observe?
5. Which method works best on credt-g.arff?