

RMIT University
School Science
COSC2110/COSC2111 Data Mining
Laboratory Week 1

Aims of this lab

- Learn how to find and access the Data Mining materials on KDrive
 - Learn how to run a classifier in Weka
-

These instructions assume that you are using a windows PC in a lab at RMIT.

1. Navigate to the KDrive
 - (a) Log in to your RMIT account. Click Computer. You should see “University K:”
 - (b) Navigate to `KDriveSEH\SCSIT\Students\Courses\COSC2111\DataMining`
 - (c) Look through the various sub folders and get a sense of the material here. In particular look in the “data” folder.
2. To start Weka: Navigate to `weka-3-9-2` and double click `weka.jar`, click explorer.
3. Click on **Open File** and load
`/KDrive/SEH/SCSIT/Students/Courses/COSC2111/DataMining/data/arff/UCI/iris.arff`.
4. Click **Classify** and select IBk from the lazy classifiers
 - (a) Click Percentage split
 - (b) Click start, what is the test accuracy?
 - (c) By default 66% of the available data is being used for training. Gradually lower this percentage. What is the effect on test accuracy? Is this what you expect?
 - (d) Go back to the default 66%split. IBk has a parameter k . What is k ? Rerun the classifier for increasing values of k . What is the effect? Is this what you expect?
5. Select OneR from the rule classifiers.
 - (a) Click Percentage split
 - (b) Click start, what is the test accuracy?
 - (c) By default 66% of the available data is being used for training. Gradually lower this percentage. What is the effect on test accuracy? Is this what you expect? How does it differ from the case of IB1?

- (d) Go back to the default 66% split. OneR has a parameter *minBucketSize*. What is *minBucketSize*? Rerun the classifier for various values of bin. What is the effect? Is this what you expect?
- 6. Experiment with other classifiers and data files.
- 7. Local copy of Weka.
 - (a) In some labs you will be able to run it from “All Programs” on your PC
 - (b) Or find it through 'Search Programs and Files'.