RMIT University
School of Science
COSC2110/COSC2111 Data Mining
Laboratory Week 3

Aims of this lab

- Learn how run numeric prediction algorithms and interpret the results.
- Learn how adjust the C and M parameters of J48.

---

1. The data files for this lab can be found at `/KDrive/SEH/SCSIT/Students/Courses/COSC2111/DataMining/data`
2. Load the file `arff/UCI/iris.arff`.
   (a) Go to the classifiers screen and select `sepallength` as the class attribute.
   (b) Select 'More Options' and 'Output Predictions'.
   (c) Run the M5P classifier with default parameters.
   (d) Examine the output. What do you think of the accuracy of the predictions?
   (e) Experiment with different values for the parameters. What is the effect on accuracy?
   (f) Experiment with ZeroR and IBK and their various parameters.
   (g) Build a table of classifier, parameter values and error. What combination gives the most accurate predictions?
   (h) Can you explain the differences in errors?
3. Repeat the previous exercise with `cpu.with.vendor.arff`
4. Load the file `soybean.arff`
   (a) Run the J48 classifier with default parmeters.
   (b) Make sure you can visualise the tree.
   (c) Experiment with different values of the C and M parameters.
   (d) Using percentage split build a table of training and test errors
   (e) Is there any overfitting?
   (f) What would you say is the best combination of parameter values?
5. Repeat the previous exercise with the `glass.arff`
6. Experiment with other classifiers and data files.