Department of Computer Engineering
University of Peradeniya

CO 544 Machine Learning and Data Mining
Lab 01

05th of March 2020

Objectives

- Exploring a data set and identifying its attributes using the tool, **Weka**.

- Selecting data mining algorithms, test options and parameters of the algorithms when constructing decision trees.

Exercises

1. Download the **hepatitis.arff** data set.

   The data set contains details of patient id, age, sex, conditions of the several measurements in the liver, chemical compound levels and type of each patient (LIVE,DIE).

2. View the data file in ARFF-viewer.

3. Start up **Weka** and import the data set.

4. What is the data type of attribute **'ALBUMIN'**?

5. How many missing values in attribute **'SPIDERS'**?

6. Build the **J4.8** decision tree using the training set and visualize the tree.

7. Build the **J4.8** decision tree using the cross validate with 10 folds and visualize the tree.

8. Note down the results in the below table.

	Correctly classified instances	Incorrectly classified instances
Training Set		
Cross validation(10 folds)		

9. Interpret the results from the '**confusion matrix**' in the classifier output.

10. Change the parameters as below and compare the results with the results of the above model with default values. **Confidence factor**: 0.5 and **Min number of folds**: 2

Notes

You should complete the exercies upto 7 and get them marked within the lab by an instructor.
Submit your answers to Exercises 8-10 as a PDF before the next lab.