

Instructions on Weka Usage for Project 2

UE 141 on Data Mining Spring 2013

Documentation can be found from: <http://www.cs.waikato.ac.nz/ml/weka/documentation.html>
Especially you can find a manual under the Weka folder after you install Weka.

You need to use Weka Explorer. The following functions will be used in Project 2:

1. Loading data: Discussed in Section 4.2.1 (Open file->choose the file to be open->Open)
2. Click the Classify tab.
3. By default, ten-fold cross validation is used to evaluate the classification accuracy. We don't need to change anything here.
4. By default, class label attribute is the last column. If this is the case for your data, you don't need to change anything. Otherwise, you can change it to the other column at the left of the panel.
5. Choose a classifier. We will use each of the following two classifiers:

weka->classifiers->trees->J48 (decision tree) or weka->classifiers->lazy->IB1 (nearest neighbor)
6. Click start.
7. After the classification is done, check the right panel for classification models and classification accuracy.

Datasets can be found at:

- The folder "data" under Weka folder on your machine
- <http://repository.seasr.org/Datasets/UCI/arff/>
- <http://archive.ics.uci.edu/ml/datasets.html> (You should store the data as *.csv file)