

Instructions on Weka Usage for Project 1

UE 141 on Data Mining Spring 2013

You can download Weka from: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
Choose from the stable versions based on your computer Operating System. If you are not sure whether your computer has Java VM installed or not, choose the one that includes Java VM.

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 1:

1. Loading data: Discussed in Section 4.2.1 (Open file->choose the file to be open->Open)
2. Discretize the data if some attributes are numeric: Discussed in Section 4.2.4 (Filter->Click Choose->Under weka-filters-unsupervised-attribute-Discretize->Apply)
3. Association analysis: Discussed in Section 4.5 (Associate-> Choose Associator -> Apriori->Change option values->Start)
Options for Association analysis:
Possible options are listed at: <http://wiki.pentaho.com/display/DATAMINING/Apriori>
LowerBoundMinSupport: Support threshold (0-1)
MinMetric: Confidence threshold (0-1)
NumRules: The number of rules to be returned. You can set a large number to return all the rules.

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)