Weka Practical III

Note:

- Briefly, regression algorithms deal with numerical attributes, while Apriori (an association rule algorithm) only accept nominal values. Regression algorithms that will be used for this week's exercises include LinearRegression, SimpleLinearRegression, Logistic (Optional) and SimpleLogistic (Optional). You need to identify a class attribute in the dataset when you want to apply any of these regression algorithms. Also Logistic algorithms deal with two class regression problems. Apriori generates association rules that do not focus on any particular attribute (no class used here).

- Read these two files if you have any weka related questions first: *How to Use Weka* **and** *WEKA Explorer User Guide*.
- Don't change any parameters until you understand them.
- Study each of the datasets by opening the .arff file in wordpad before you apply any algorithm to it. Metadata is normally included at the beginning in the file. For example, for each of the datasets for regression exercises, the purpose of the dataset, such as the class attribute, and the meaning of each attribute, etc. are explained.

Session 1

Use LinearRegression and SimpleLinearRegression (if applicable) to analyse the following datasets, which can be downloaded from Moodle:

- cpu.arff and cpu.with.vendor.arff
- bodyfat.arff
- breastTumor.arff
- cal_housing.arff
- housing.arff

Optional: Use Logistic and SimpleLogistic to analyse weather.arff.

Session 2

Use Aprioi to analyse the following datasets, which are included in Weka Practical 1 and II:

- weather.nominal.arff
- contact-lenses.arff
- titanic.arff
- mushroom.arff
- lifestyle (optional)

Note: There are 36 attributes in the lifestyle data. You need to select a smaller group of attributes for analysing.