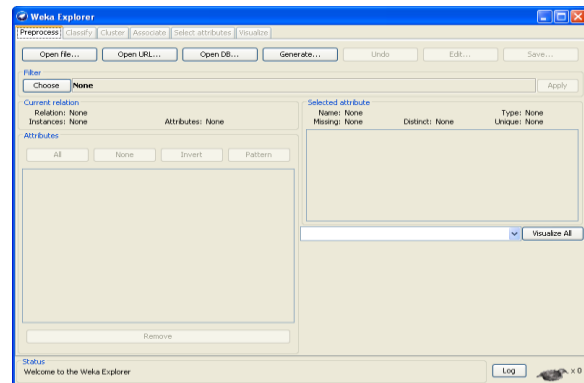
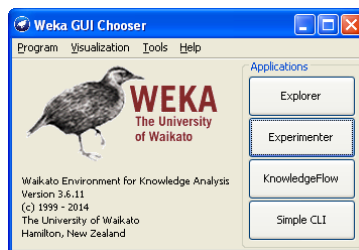


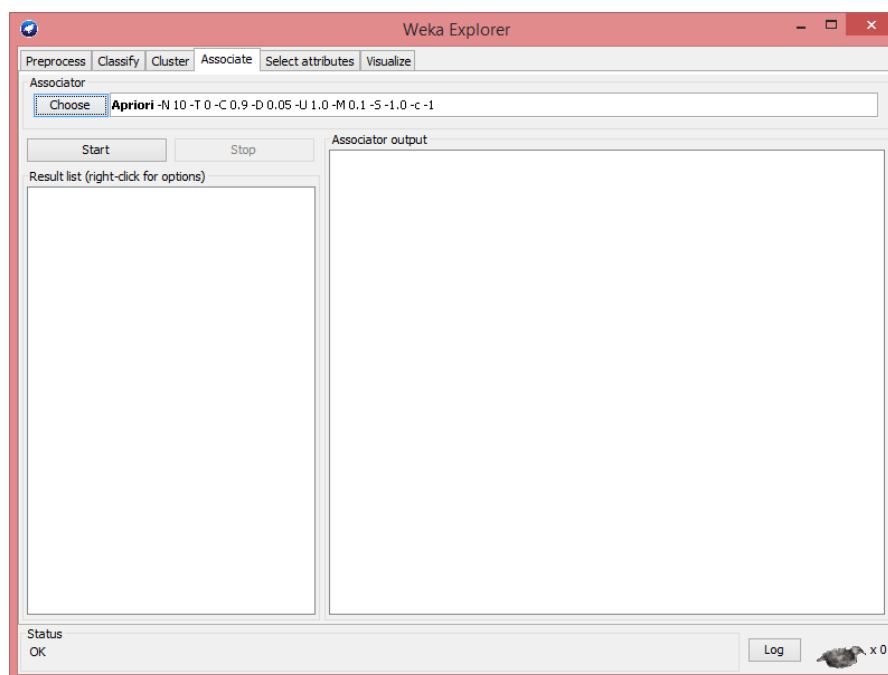
Lab Exercise Two

Mining Association Rule with WEKA Explorer

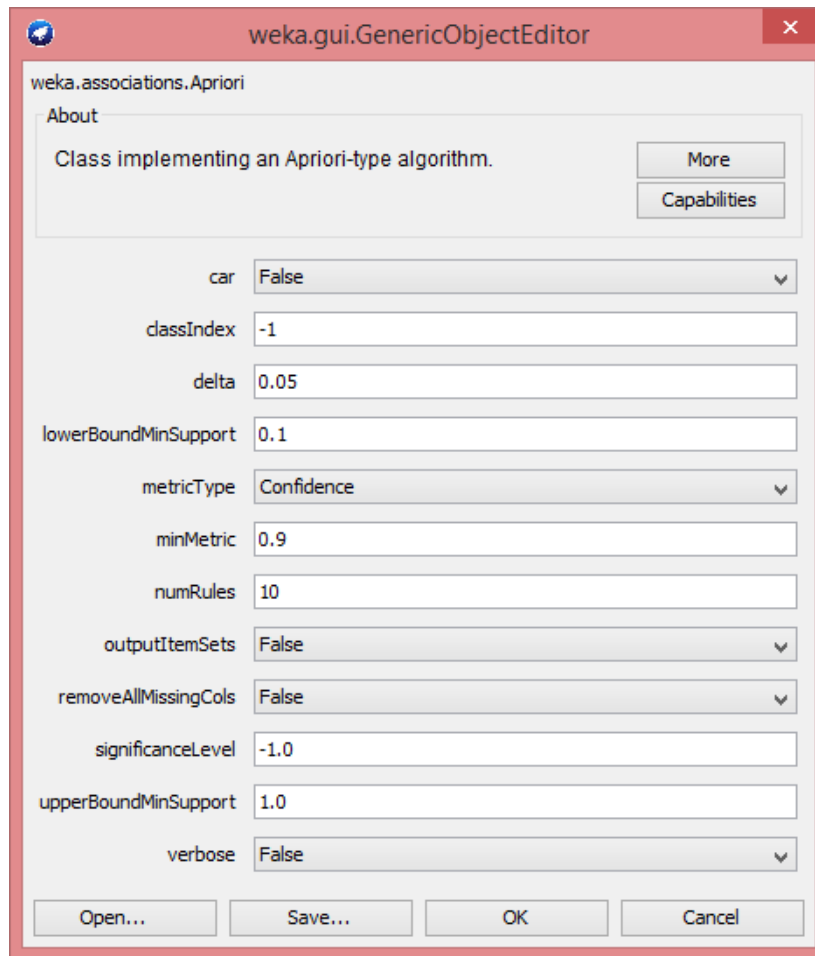
1. Open a terminal window from the left bar. Go to directory `/opt/weka-3-6-13`, then type command :
java -jar weka.jar.
2. Fire up WEKA to get the GUI Chooser panel. Select Explorer from the four choices on the right side.



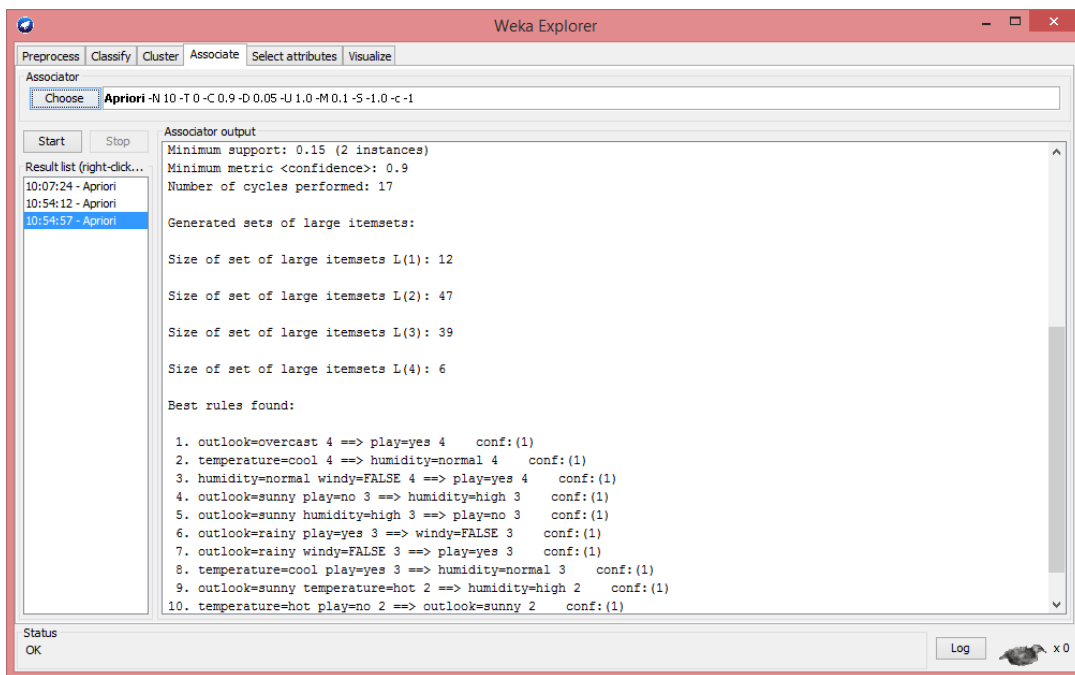
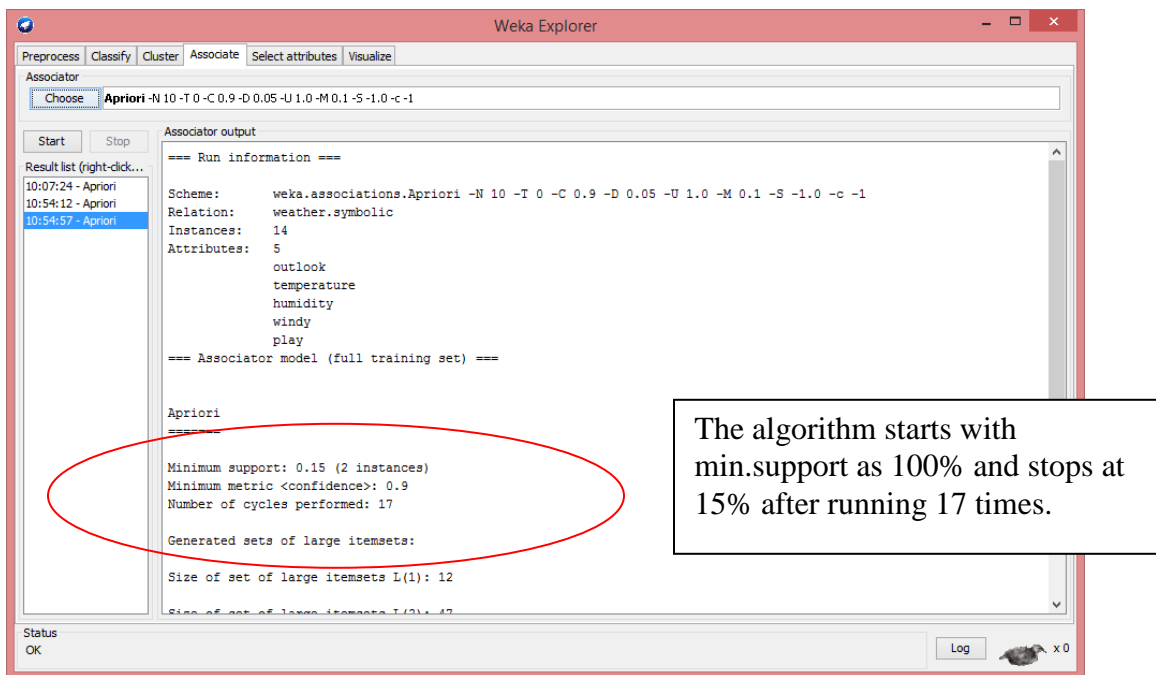
3. To get a feel for how to apply Apriori, start by mining association rules from the *weather.nominal.arff* data set of Lab One. Note that Apriori algorithm expects data that is purely nominal: If present, numeric attributes must be discretized first.
4. Click Associate Tab on top of the window.



5. Left click the field of Associator, choose Show Property from the drop down list. The property window of Apriori opens:



6. Weka runs an Apriori-type algorithm to find association rules, but this algorithm is not exact the same one as we discussed in class.
 - a. The min. support is not fixed. This algorithm starts with min. support as **upperBoundMinSupport** (default 1.0 = 100%), iteratively decrease it by **delta** (default 0.05 = 5%). Note that **upperBoundMinSupport** is decreased by delta *before* the basic *Apriori* algorithm is run for the first time.
 - b. The algorithm stops when **lowerBoundMinSupport** (default 0.1 = 10%) is reached, or required number of rules – **numRules** (default value 10) have been generated.
 - c. Rules generated are ranked by **metricType** (default *Confidence*). Only rules with score higher than **minMetric** (default 0.9 for *Confidence*) are considered and delivered as the output.
 - d. If you choose to show the all frequent itemsets found, **outputItemSets** should be set as *True*.
7. Click **Start** button on the left of the window, the algorithm begins to run. The output is showing in the right window.



- You could re-run Apriori algorithm by selecting different parameters, such as *lowerBoundMinSupport*, *minMetric* (min. confidence level), and different evaluation metric (confidence vs. lift), and so on.

9. Now consider a real-world dataset, *vote.arff*, which gives the votes of 435 U.S. congressmen on 16 key issues gathered in the mid-1980s, and also includes their party affiliation as a binary attribute. Let's use association-rule mining can also be applied to this data to seek interesting associations.

We are on **Preprocess** now. Click the **Open file** button to bring up a standard dialog through which you can select a file. Choose the *vote.arff* file. To see the original dataset, click the **Edit** button, a viewer window opens with dataset loaded. This is a purely nominal dataset with some missing values (corresponding to abstentions).

Exercise 1. Run *Apriori* on this data with default settings. Comment on the rules that are generated. Several of them are quite similar. How are their support and confidence values related?

Exercise 2. It is interesting to see that none of the rules in the default output involve *Class = republican*. Why do you think that is?

10. Let's run *Apriori* on another real-world dataset. We are on **Preprocess** now. Click the **Open file** button to bring up a standard dialog through which you can select a file. Choose the *supermarket.arff* file. To see the original dataset, click the **Edit** button, a viewer window opens with dataset loaded. To do market basket analysis in Weka, each transaction is coded as an instance of which the attributes represent the items in the store. Each attribute has only one value: If a particular transaction does not contain it (i.e., the customer did not buy that item), this is coded as a missing value.

Exercise. Experiment with *Apriori* and investigate the effect of the various parameters described before. Write a brief report on the main findings of your investigation.