Association rules: computer practice

1) Basic manipulation

Let us consider the table below with obvious interpretation (each line is a customer and 1 2 3 means that the corresponding customer bought item 1 item 2 and item3.
Customer table:

1 2 3
1 4 5
2 3 4
1 2 3 4
2 3
1 2 4
4 5
1 2 3 4
3 4 5
1 2 3

1. Considering the association rule 4 <- 5, provide the support level and the confidence level of this rule.
2. Extract at least a rule with less than 20% as support level. Compute its confidence level.

3. The *lift* value of an association rule is the factor by which the confidence exceeds the expected confidence. It is determined by dividing the confidence of the rule by the support of the rule head. Compute the lift for the following rules:

- 3 <- 1 2
- 2 <- 1 3
- 2 <- 3

2) Weka

Install weka and download the **weather** dataset. Use weka to extract association rules from the dataset. Try different parameters and try different algorithms (APriori, Tertius). Analyze the differences.
Get information from the internet about Tertius algorithm.
In term of complexity, investigate the complexity of APriori (in term of the size of the initial dataset).

3) UCI

University of California at Irvine repository is one of the most famous dataset repository for machine learning.
Go to the nominal data folder, choose one or 2 datasets and download them.
Investigate the datasets via Weka and association rules.

4) Internet data

We want to analyze raw data coming from the Internet (nominal data). Dowload such a dataset, then pre-process this dataset to make it suitable for Weka. Extract association rules.

5) Internet data

Get the **apriori** executable from the internet and build up a simple web site where a user can upload a data file (csv format for instance or a text file according to the apriori requirement) then get the result of the apriori search in the browser. The user should the facilities to choose his/her own confidence, support level at least (i.e I want the rules with support greater than 30% and confidence greater than 85%).