HW 2: Association Rules

1. Apriori by hand

Apply the Apriori algorithm on the transactions of the table illustrated below, with minimum support, s = 40% and minimum confidence, c = 70%. Specify the technique you use for itemset candidate generation and show the candidate and the frequent itemsets at each step of the algorithm. Give all the frequent itemsets and the corresponding association rules sorted by confidence. For the qualifying rules (rules that exceed the minimum confidence), also compute their lift. Discuss which you consider to be the strongest rules overall and why.

Transaction ID	Items
T1	beer, eggs, diapers, bread, cheese
T2	eggs, diapers
T3	beer, eggs, milk
T4	eggs, beer, cheese
T5	beer, milk, diapers

2. Apriori with Weka

Consider the data file house-votes-84.csv that is based on the 1984 United Stated Congressional Voting Records found at the UCI Machine Learning Repository (https://doi.org/10.24432/C5C01P). The data contain 16 categorical attributes describing the votes of the Congressmen on 16 key issues, and one class attribute depicting whether the Congressman is of the Republican or Democratic party. For the 16 attributes, consider that they are binary attributes, with values 1 (voted yes), and 0 (voted no), while there are also some unknown values (?). Use WEKA to find association rules for this dataset. In particular:

1) Create an appropriate .arff file.

2) Replace missing values (unknown values) with the most frequent value in each category (i.e., for each attribute).

3) Use Apriori to mine association rules with minimum support, *s=0.5* and minimum confidence, *c=0.9.*

4) Use Apriori to mine association rules with minimum support, *s=0.5* and minimum lift, *l=1.5.*

5) Compare the results of 3 and 4 for the top-20 rules (i.e., how similar are they, which are the strongest rules overall, etc)

6) Use Apriori to mine association rules with appropriate support and confidence so as to derive association rules that could be used for classifying a voter as a Republican or Democrat. Discuss your results.

Submit a report with the solution of problem 1 and all comments for problem 2.

For problem 2 also submit the corresponding .arff file you used, as well as result files for queries 3, 4 and 6 (save the weka output as .txt files).