Assignment 5

Due: Dec. 7

1. (25 points) Consider a dataset for frequent set mining as in the following table where we have 6 binary features and each row represents a transaction.

0 0 1 0 1 0
0 1 1 1 0 1
1 0 0 0 1 0
1 1 1 0 0 0
0 0 0 1 0 0
1 0 0 1 0 1
0 0 1 1 1 1
1 0 1 0 1 0
1 0 0 1 0 0
0 1 1 0 0 1

   (a) Illustrate the first three levels of the Apriori algorithm (set sizes 1, 2 and 3) for support threshold of 3 transactions, by identifying candidate sets and calculating their support. What are the maximal frequent sets discovered in the first 3 levels?

   (b) Pick one of the maximal sets and check if any of its subsets are association rules with frequency at least 0.3 and confidence at least 0.6. Pleas explain your answer and show your work.

2. (25 points) Given the following transaction database, let the min_support = 2, answer the following questions.

TID	Items
1	{a,b,e}
2	{a,b,c,d}
3	{a,c,d}
4	{a,c,e}
5	{b,c,f}
6	{a}
7	{a,b,c}
8	{b,d,e}
9	{a,c}
10	{a,b,d,e}

(a) Construct FP-tree from the transaction database and draw it here.

(b) Show d's conditional pattern base (projected database), d's conditional FP-tree, and find frequent patterns based on d's conditional FP-tree.

3. (25 points) In the GSP algorithm, suppose we have the length-3 frequent pattern set L_3 as follows:

$$
\begin{array}{l}
< \{2\}\ \{3\}\ \{4\} > \\
< \{2\ 5\}\ \{3\} > \\
< \{3\}\ \{4\}\ \{5\} > \\
< \{1\}\ \{2\}\ \{3\} > \\
< \{1\}\ \{2\ 5\} > \\
< \{1\}\ \{5\}\ \{3\} > \\
< \{5\}\ \{3\ 4\} >
\end{array}
$$

Generate length-4 candidates set C_4 and frequent pattern set L_4. Show your work by writing down the details of the join and prune steps.

4. (25 points) For the following two time series:

$$
\begin{aligned}
X &= [39\ 44\ 43\ 39\ 46\ 38\ 39\ 43] \\
Y &= [37\ 44\ 41\ 44\ 39\ 39\ 39\ 40]
\end{aligned}
$$

Calculate the DTW distance between X and Y and point out the optimal warping path. (The local cost function is defined as the absolute difference of the two values, e.g., $c(x_1, y_1) = d(39, 37) = 2$)