Assignment 5

Due: May 6

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) Table 1 shows how many transactions containing beer and/or nuts among 10000 transactions. Answer the following questions.

	Beer	No Beer	Total
Nuts	50	800	850
No Nuts	150	9000	9150
Total	200	9800	10000

Table 1: Transactions Involving Beer and Nuts.

   (a) How many possible association rules can be generated based on Table 1?

   (b) Calculate support, confidence, lift for each of the rules in (a).

   (c) What are your conclusions of the relationship between buying/not buying beer and buying/not buying nuts, based on the above measures?

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

$< \{2\}\ \{3\}\ \{4\} >$
$< \{2\ 5\}\ \{3\} >$
$< \{3\}\ \{4\}\ \{5\} >$
$< \{1\}\ \{2\}\ \{3\} >$
$< \{1\}\ \{2\ 5\} >$
$< \{1\}\ \{5\}\ \{3\} >$
$< \{5\}\ \{3\ 4\} >$

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

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

$$X = [32, 36, 27, 37, 35, 40, 34, 33, 25, 29]$$
$$Y = [31, 32, 32, 30, 37, 39, 29, 34, 25, 26]$$

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., $c(x_1, y_1) = d(32, 31) = 1$)