

UCLA CS 145 Homework #4

DUE DATE: Wednesday, 11/29/2017 11:59 PM

Note

- You are expected to submit your answers as a report named `CS145_HW4_UID.pdf` to CCLE.
- Copying and sharing of homework are NOT allowed. But you can discuss general challenges and ideas with others. *Suspicious cases will be reported to The Office of the Dean of Students.*
- Late submission policy: you will get a late factor $\mathbb{1}(t \leq 24) \times e^{(-\ln(2)/12) \times t}$, if you are t hours late.

1 Frequent Pattern Mining for Set Data

Given a transaction database shown in Table 1, answer the following questions. Note that the parameter `min_support` is set as 2.

- Find all the frequent patterns using Apriori Algorithm. Details of the procedure are expected.
- Construct and draw the FP-tree of the transaction database.
- For the item d , show its conditional pattern base (projected database) and conditional FP-tree.
- Find frequent patterns based on d 's conditional FP-tree.

Table 1: The transaction database for the question 1.

TID	Items
1	b, c, j
2	a, b, d
3	a, c
4	b, d
5	a, b, c, e
6	b, c, k
7	a, c
8	a, b, e, i
9	b, d
10	a, b, c, d

2 Correlation Analysis

Table 2 shows how many transactions containing beer and/or nuts among 10000 transactions. Answer the following questions based on Table 2.

- Calculate **confidence**, χ^2 , **lift**, and **all_confidence** between buying beer and buying nuts.
- What are your conclusions of the relationship between buying beer and buying nuts, based on the above measures?

Table 2: Contingency table for question 2.

	Beer	No Beer	Total
Nuts	150	700	850
No Nuts	350	8800	9150
Total	500	9500	10000

3 Sequential Pattern Mining (GSP Algorithm)

- For a sequence $s = \langle ab(cd)(ef) \rangle$, how many events or elements does it contain? What is the length of s ? How many non-empty subsequences does s contain?
- Suppose we have $L_3 = \{ \langle (ac)e \rangle, \langle b(cd) \rangle, \langle bce \rangle, \langle a(cd) \rangle, \langle (ab)d \rangle, \langle (ab)c \rangle \}$ as the frequent 3-sequences, write down all the candidate 4-sequences C_4 with the details of the join and pruning steps.

4 Dynamic Time Warping (DTW)

Suppose that we have two sequences S_1 and S_2 as follows:

$$S_1 = \langle 1, 2, 5, 3, 2, 1, 7 \rangle$$

$$S_2 = \langle 2, 3, 2, 1, 7, 4, 3, 0, 2, 5 \rangle$$

Compute the distance between two sequences according to the dynamic time warping algorithm. The local distance (cost) between two items is defined as $c(x, y) = |x - y|$. Note that you need to write down the detailed procedure to determine the distance.