CSE 40647/60647 Data Science (Spring 2018)
Lecture 21: Frequent Pattern Mining: Apriori

Concepts in Frequent itemset mining

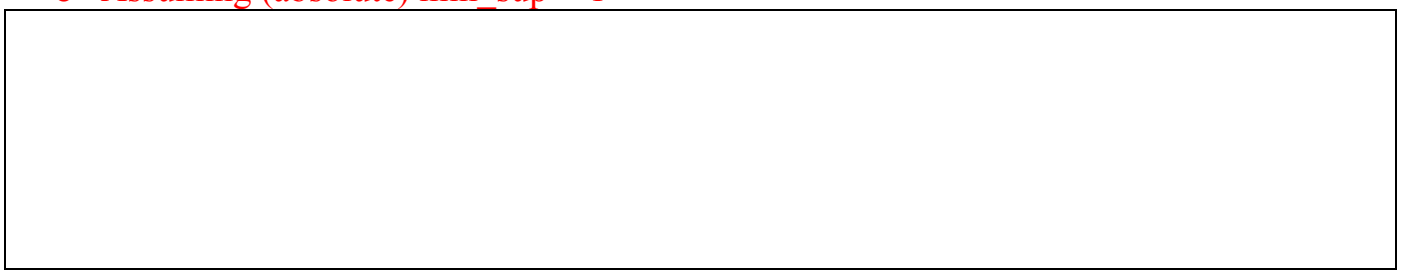
- Itemset, k-itemset, absolute support, relative support, frequent itemset

- Association rules, support, confidence

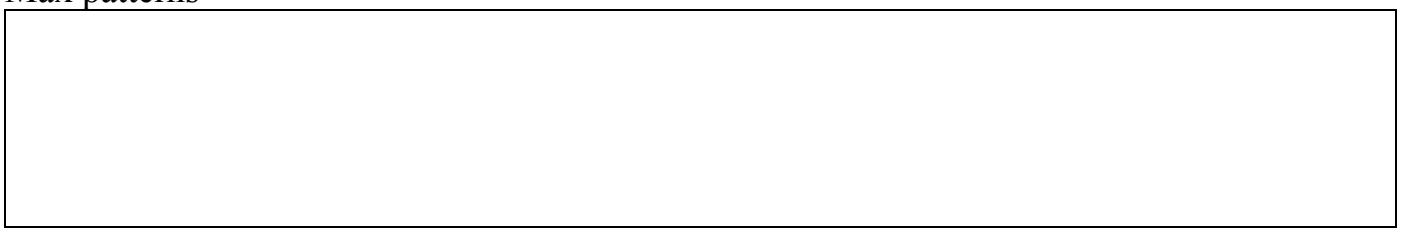
- Q1: How many frequent itemsets does the following transaction database (TDB) contain?
  - T1: $\{a_1, \ldots, a_{50}\}$; T2: $\{a_1, \ldots, a_{100}\}$
  - Assuming (absolute) min_sup = 1

- Closed patterns

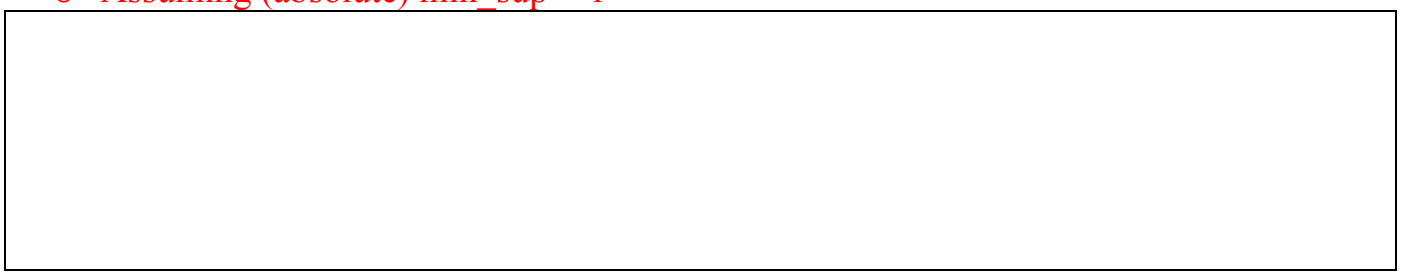
- Q2: How many closed patterns does TDB contain? What are they?
  - T1: $\{a_1, \ldots, a_{50}\}$; T2: $\{a_1, \ldots, a_{100}\}$
  - Assuming (absolute) min_sup = 1



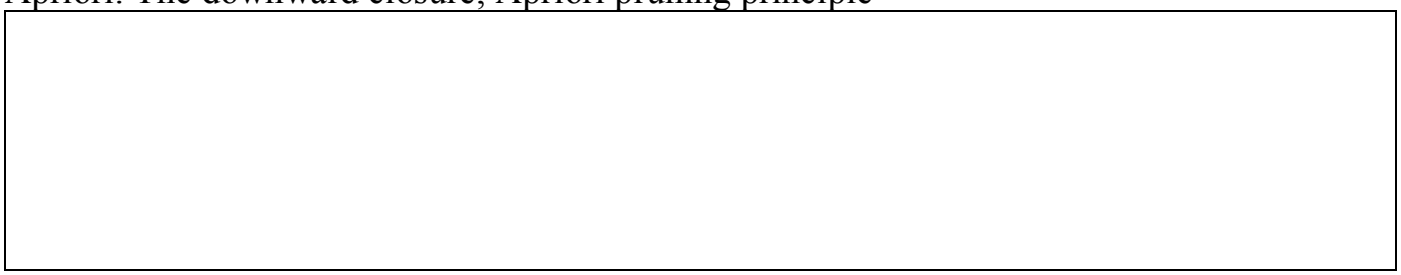
- Max patterns



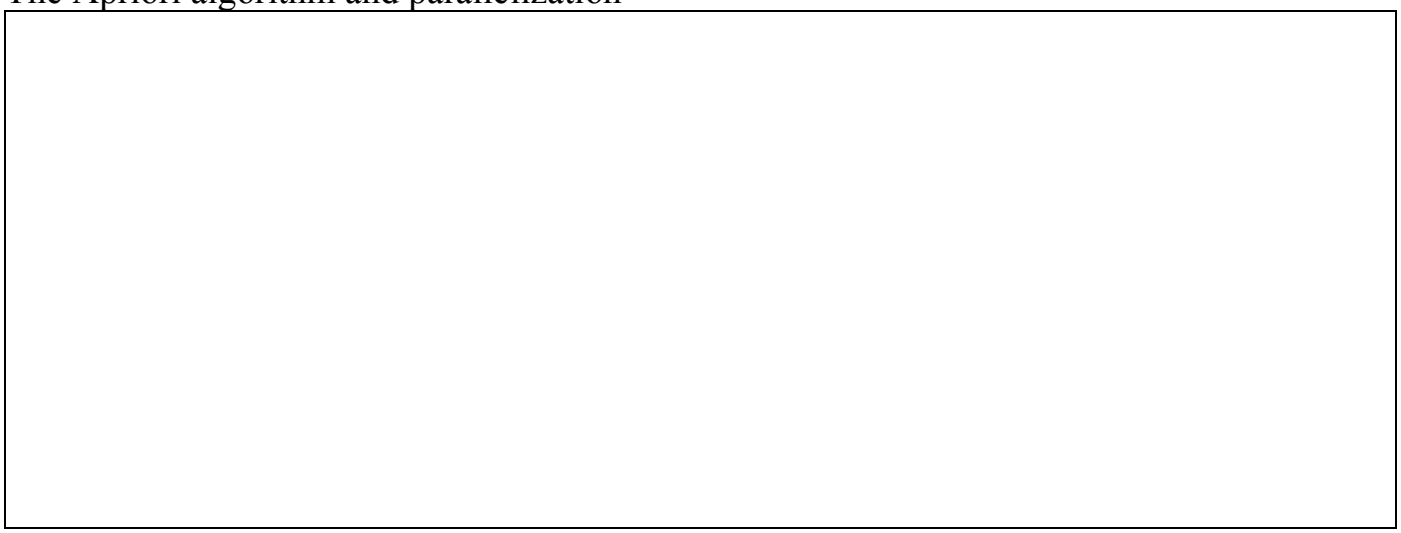
- Q3: How many max patterns does TDB contain?
  - T1: $\{a_1, \ldots, a_{50}\}$; T2: $\{a_1, \ldots, a_{100}\}$
  - Assuming (absolute) min_sup = 1



- Apriori: The downward closure; Apriori pruning principle



- The Apriori algorithm and parallelization



Name (NetID):

Given a transaction database: (you can simplify the item names as their first letter)

Transaction ID	Items Bought
T1	{Mango, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T2	{Doll, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T3	{Mango, Apple, Key-chain, Eggs}
T4	{Mango, Umbrella, Corn, Key-chain, Yo-yo}
T5	{Corn, Onion, Onion, Key-chain, Ice-cream, Eggs}

Answer the following questions if min_sup = 60%:

(1) Use Apriori to find all the frequent itemsets and their support.

(2) List the closed patterns and their absolute support.

(3) List the max patterns and their absolute support.