Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining — Assignment 2

Due: January 25, 2024(100 points)

YOUR NAME
YOUR GIT USERNAME
YOUR E-MAIL

Frequent Itemsets

Consider the following set of frequent 3-itemsets:

```
{1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4},
{2, 3, 4}, {2, 3, 5}, {3, 4, 5}.
```

Assume that there are only five items in the data set. This question was taken from Tan et al., which may help in reviewing Candidate Generation.

1. List all candidate 4-itemsets obtained by a candidate generation procedure using the $F_{k-1} \times F_1$ merging strategy.

2. List all candidate 4-itemsets obtained by the candidate generation procedure in A Priori, using $F_{k-1} \times F_{k-1}$.

3. List all candidate 4-itemsets that survive the candidate pruning step of the Apriori algorithm.

Association Rules

Consider the following table for question 4:

Transaction ID	Items
1	**{Beer, Diapers}**
2	**{Milk, Diapers, Bread, Butter}**
3	**{Milk, Diapers, Cookies}**
4	**{Bread, Butter, Cookies}**
5	**{Milk, Beer, Diapers, Eggs}**
6	**{Beer, Cookies, Diapers}**
7	**{Milk, Diapers, Bread, Butter}**
8	**{Bread, Butter, Diapers}**
9	**{Bread, Butter, Milk}**
10	**{Beer, Butter, Cookies}**

4. a) What is the maximum number of association rules that can be extracted from this data (including rules that have zero support)?
   b) What is the confidence of the rule {Milk, Diapers} $\Rightarrow$ {Butter}?
   c) What is the support for the rule {Milk, Diapers} $\Rightarrow$ {Butter}?

5. `True` or `False` with an explanation: Given that {a,b,c,d} is a frequent itemset, {a,b} is always a frequent itemset.

6. `True` or `False` with an explanation: Given that {a,b}, {b,c} and {a,c} are frequent itemsets, {a,b,c} is always frequent.

7. `True` or `False` with an explanation: Given that the support of {a,b} is 20 and the support of {b,c} is 30, the support of {b} is larger than 20 but smaller than 30.

8. `True` or `False` with an explanation: In a dataset that has 5 items, the maximum number of size-2 frequent itemsets that can be extracted (assuming minsup > 0) is 20.

9. Draw the itemset lattice for the set of unique items $\mathcal{I} = \{a, b, c\}$.