Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining | Assignment 2

Due: February 25, 2023(100 points)

Yichen Sun
https://github.com/LAnselet/cs6220-datamining

Frequent Itemsets

Question 1

F_{k-1} x F_1 = F_3 x F_1 = {1, 2, 3, 4}, {1, 2, 3, 5}, {1, 2, 4, 5}, {1, 3, 4, 5}, {2, 3, 4, 5}

Question 2

F_{k-1} x F_k - 1 = F_3 x F_3 = {1, 2, 3, 4}, {1, 2, 3, 5}, {1, 2, 4, 5}, {2, 3, 4, 5}

Question 3

C_4 after prune = {1, 2, 3, 4}

Association Rules

Question 4

a. $R = 3^n - 2^{n+1} + 1 = 3^7 - 2^8 + 1 = 1932$
b. $\frac{2}{4} = 0.5$
c. $\frac{2}{10} = 0.2$

Question 5

True. As Apriori principle illustrated. If {a, b, c} if a frequent itemset, then all subsets of this item set are frequent.

Question 6

False. {a, b, c} may not even exist.

Question 7

False. The support of {b} could be 50 without overlapping.

Question 8

False. The maximum can potentially generate up to $2^k - 1$ frequent itemsets which will be $2^5 - 1$ itemsets.

Question 9