CZ4032 Data Analytics and Mining

Tutorial for Week 7: association rules

Q1 The following is a pseudocode of Apriori algorithm

1) Generate frequent itemsets of length 1
2) Repeat until no new frequent itemsets are identified
    a) Generate length (k+1) candidate itemsets from length k frequent itemsets
    b) Prune candidate itemsets containing subsets of length k that are infrequent
    c) Count the support of each candidate by scanning the DB
    d) Eliminate candidates that are infrequent, leaving only those that are frequent

Explain 2 b) with an example.

Q2 Illustrating Apriori Principle with an example.

Q3 In the lecture, we introduce how to generate length (k+1) candidate itemsets from length k frequent itemsets. Explain this with an example. Can give another way of generating candidates?

Q4 A dataset with 4 records

1: a, c,d
2, b, c,e
3, a, b, c, e
4, b, e,

Suppose minimum support is 2. Find frequent itemsets step-by-step.

Q5 Suppose {B,C,D} is a frequent itemset. Enumerate the candidate rules:

Q6 Consider the observation: If A,B,C→D is below confidence, so is A,B→C,D. Can we design an efficient order of generating rules based on the observation?