

Problem 1. Consider the following dataset. Extract all the frequent itemsets with a minimum support of 3 and a minimum confidence equal to 0.8.

A	B	C	D
0	0	0	1
0	1	1	0
1	0	1	1
1	1	0	0
0	0	0	1
1	0	0	0
1	0	0	0
1	1	1	1
1	0	1	1

Problem 2. Consider the following market-basket data represented in a two-attribute table (where T# is the transaction identifier). Specify all of the association rules that can be deduced from this data with **Support > 0.3** and **Confidence > 0.5**. To limit your search, only consider association rules that have exactly one item on the left-hand side and one item on the right-hand side.

T#	item
1	cookies
1	milk
2	beer
2	pretzels
2	cookies
2	eggs
3	beer
3	pretzels
4	beer
4	cookies
4	milk
5	beer
5	cookies

Problem 3. A database has four transactions. Let min_sup = 60% and min_conf = 80%. Find all frequent items using Apriori.

TID	Date	Items bought
T100	10/15/05	{K, A, D, B}
T200	10/15/05	{D, A, C, E, B}
T300	10/19/05	{C, A, B, E}
T400	10/22/05	{B, A, D}

Problem 4. Given the following dataset, let the minimum support threshold be 60% and the minimum confidence threshold be 80%. Find all frequent itemsets and list the strong association rules.

TID	items
T ₁₀₀	{M, O, N, K, E, Y}
T ₂₀₀	{D, O, N, K, E, Y}
T ₃₀₀	{M, A, K, E}
T ₄₀₀	{M, U, C, K, Y}
T ₄₀₀	{C, O, K, I, E}

Problem 5. You have run the a-priori algorithm to find association rules in a grocery store transaction database. It takes an unexpectedly long time to complete. On completion, the following is one (of many) rules: "<milk, butter, cheese, bread, flour, sugar, salt, chocolate, apples> => vanilla"
Based on seeing the above rule, you should be able to make a good guess as to why the algorithm took a long time. Explain why.