**6.3** The Apriori algorithm makes use of prior knowledge of subset support properties.

(a) Prove that all nonempty subsets of a frequent itemset must also be frequent.

If the set A contains only frequent items, then a subset of A will contain only frequent items, as well. This is because there are no infrequent items in the lager set A, nor are there any infrequent items in its subset. In fact, any non-empty subset of A will contain frequent items, because the subset inherits the properties of its superset.

So if we have set A = {<milk, bread, candy>, <milk, candy>, <bread, milk>}, and if we know that for all elements x in the set A x are frequent items, then a subset of A {<bread, milk>} will still be frequent items.

(b) Prove that the support of any nonempty subset s0 of itemset s must be at least as great as the support of s.

Let’s say our itemset s contains {pants, vests, ties, loafers, blouses}. Then say we take a proper subset s0 = {ties, loafers}. The support would be the co-occurrence of {ties, loafers} out of all of the transactions under analysis. The support for subset s0 cannot be less than the support for s, because there are fewer items in the subset s0. Therefore, the number of items compared for this association rule are fewer, and the possible combinations are also fewer.

(c) Given frequent itemset *1* and subset *s* of *l*, prove that the confidence of the rule “*s ⇒ (l - s′)”* cannot be more than the confidence of “*s ⇒ (l - s),”* where *s′* is a subset of *s*.

*s′* is a subset of *s*, and *s* is a subset of *l*. So *(l – s)* must be ≤ *(l – s′)*, since there are possibly more elements in s than there are in s*′.*The association rule of confidence for *s ⇒ (l - s) =* P(*(l - s)*|*s*)  
which is the same as

P which is ≥ P

(d) A partitioning variation of Apriori subdivides the transactions of a database D into n nonoverlapping partitions. Prove that any itemset that is frequent in D must be frequent in at least one partition of D.

**6.6** A database has five transactions. Let min sup D 60% and min conf D 80%.

(a) Find all frequent itemsets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.

(b) List all the strong association rules (with support s and confidence c) matching the following metarule, where X is a variable representing customers, and itemi denotes variables representing items (e.g., “A,” “B,”): 8x 2 transaction, buys.X, item1/^buys.X, item2/)buys.X, item3/ [s, c]

**6.8** A database has four transactions. Let min sup D 60% and min conf D 80%.

(a) At the granularity of item category (e.g., itemi could be “Milk”), for the rule template, 8X 2 transaction, buys.X, item1/^buys.X, item2/)buys.X, item3/ [s, c], list the frequent k-itemset for the largest k, and all the strong association rules (with their support s and confidence c) containing the frequent k-itemset for the largest k.

(b) At the granularity of brand-item category (e.g., itemi could be “Sunset-Milk”), for the rule template, 8X 2 customer, buys.X, item1/^buys.X, item2/)buys.X, item3/, list the frequent k-itemset for the largest k (but do not print any rules).

**6.13** Give a short example to show that items in a strong association rule actually may be negatively correlated.

**6.14** The following contingency table summarizes supermarket transaction data, where hot dogs refers to the transactions containing hot dogs, hot dogs refers to the transactions that do not contain hot dogs, hamburgers refers to the transactions containing hamburgers, and hamburgers refers to the transactions that do not contain hamburgers.

(a) Suppose that the association rule “hot dogs 🡪 hamburgers” is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong?

(b) Based on the given data, is the purchase of hot dogs independent of the purchase of hamburgers? If not, what kind of correlation relationship exists between the two?

(c) Compare the use of the all confidence, max confidence, Kulczynski, and cosine measures with lift and correlation on the given data.