CHAPTER 28: DATA MINING CONCEPTS

28.14 Apply the Apriori algorithm to the following data set:

Trans ID	Items Purchased
101	milk, bread, eggs
102	milk, juice
103	juice, butter
104	milk, bread, eggs
105	coffee, eggs
106	coffee
107	coffee, juice
108	milk, bread, cookies, eggs
109	cookies, butter
110	milk, bread

The set of items is {milk, bread, cookies, eggs, butter, coffee, juice}. Use 0.2 for the minimum support value.

28.15 Show two rules that have a confidence of 0.7 or greater for an itemset containing three items from Exercise 14.

27.20 Consider the following set of two-dimensional records:

RID	Dimension 1	Dimension 2
1	8	4
2	5	4
3	2	4
4	2	6
5	2	8
6	8	6

Also consider two different clustering schemes: (1) where Cluster 1 contains records {1, 2, 3} and Cluster 2 contains records {4, 5, 6} and (2) where Cluster 1 contains records {1, 6} and Cluster 2 contains records {2, 3, 4, 5}. Which scheme is better and why?

28.21 Use the K-means algorithm to cluster the data from Exercise 28.20. We can use a value of 3 for K and can assume that the records with RIDs 1, 3, and 5 are used for the initial cluster centroids (means).