DAT630 Fall 2017
2017-10-06
Solutions for Exercises on Frequent Itemset Mining
Vinay Setty (vinay.j.setty@uis.no)
Can be downloaded form vinaysetty.net (under teaching)

Exercise 1) Association rule mining

A database has four transactions. Let *min_sup* = 60% and *min_conf* = 80%.

TID	date	items_bough
T1	10-15-99	{K, A, D, B}
T2	10-15-99	{D, A, C, E, B}
T3	10-19-99	{C, A, B, E}
T4	10-22-99	{B, A, D}

a) Find all frequent itemsets using the Apriori algorithm.

frequent Itemsets L(1):
A 4
B 4
D 3

Size of set of frequent itemsets L(2): 3

frequent Itemsets L(2):
A B 4
A D 3
B D 3

Size of set of frequent itemsets L(3): 1

frequent Itemsets L(3):
A B D 3

b) List all of the strong association rules (with support s and confidence c) matching the following metarule, where *X* is a variable representing customers, and $item_i$ denotes variables representing items (e.g. "A", "B", etc.):
for all $X \in$ *transaction*, $buys(X, item_1) \wedge buys(X, item_2) \Rightarrow buys(X, item_3)$ [s,c]

Best rules found:

1. B C ==> A
2. A C ==> B
3. B E ==> A

Exercise 2)

Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if i divides b with no remainder. Thus, item 1 is in all the baskets, item 2 is in all fifty of the even-numbered baskets, and so on. Basket 12 consists of items {1, 2, 3, 4, 6, 12}, since these are all the integers that divide 12. Answer the following questions:

(a)If the support threshold is 5, which items are frequent?

Items 1 to 20 occur in at least 5 baskets

(b) If the support threshold is 5, which pairs of items are frequent?

Pairs {1,[2-20]}
2, multiples of 2 which are <= 20
3, multiples of 3 which are <= 20
All other pairs only occur < 5 times