

DAT630 Fall 2017

2017-10-03

Exercises on Frequent Itemset Mining

Vinay Setty (vinay.j.setty@uis.no)

Exercise 1) Association rule mining

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

TID	date	items_bought
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}

- Find all frequent itemsets using the Apriori algorithm.
- 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 $X \in transaction$, $buys(X, item_1) \wedge buys(X, item_2) \Rightarrow buys(X, item_3) [s,c]$

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:

- If the support threshold is 5, which items are frequent?
- If the support threshold is 5, which pairs of items are frequent?