Data Mining Exercises

1 Apriori and Eclat

$DB_1 =$

Tid	transaction
1	ABC
2	BC
3	BC
4	ACD
5	ABD
6	ABE

$DB_2 =$

Tid	transaction
1	ABC
2	BCDE
3	CD
4	ABD
5	ABC

Question 1 : Do a simulation of Apriori algorithm on DB_1 with minsup = 2. For each iteration of the algorithm, give the set of candidate itemsets C_i, their supports and the set F_i of frequent itemsets.

Question 2 : Find all the association rules in DB_2 with minsup = 2 and minconf = 10%.

Question 3 : For each of the following sets F, find a corresponding database and a value of minsup such that F is the set of frequent itemsets.

1. $F_1 = \{\emptyset, A, B, C, AC\}$
2. $F_2 = \{\emptyset, A, B, C, AC, AB, BC, ABC\}$
3. $F_3 = \{\emptyset, A, B, C, AC, BC\}$

Question 4 : If the order on items is the alphabetical order, in which order are the different projected databases considered by Eclat (on each database)?

Question 5 : Do a simulation of Eclat algorithm on DB_1 (using the optimized items ordering).

2 Constraints

Question 6 : We consider the constraint $C(X) = (|X| \leq 3)$ (size less than 3). Is it monotonic or anti-monotonic (give a justification) ?

Question 7 : Explain how we can extract all frequent itemsets satisfying this constraint C. What are the different possibilities ?

3 Constraints

$DB_3 =$

Tid	a	b	c	d	e	f	g
1	0	0	0	1	1	0	0
2	1	1	0	1	0	1	1
3	0	0	0	1	0	0	0
4	1	1	1	1	1	1	0
5	1	0	0	0	1	1	0
6	0	1	0	0	0	0	1

Question 8 : We want to compute the itemsets satisfying the two constraints: $C_1(X) = (sup(X) \geq 3)$ (support greater than 3) and $C_2(X) = (|X| \geq 2)$ (size greater than 2). For each of these constraints, explain if it is monotonic or anti-monotonic (give a justification).

Question 9 : how can we use these constraints to remove some of the rows/columns of the database DB_3 ? Which row and which column can be removed in the database and why?

Question 10 : Finally, in this database, what are the itemsets that satisfy both constraints?