University of Southern Denmark
IMADA
Arthur Zimek

DM868/DM870/DS804: Data Mining and Machine Learning

Spring term 2023

Exercise 3: Closed Frequent Itemsets, Apriori

Exercise 3-1 The monotonicity of confidence (1 point)

Theorem 2.1 in the Lecture states:

Given:

- itemset X
- $Y \subset X, Y \neq \emptyset$

If $\text{conf}(Y \Rightarrow (X \setminus Y)) < c$, then $\forall Y' \subset Y$:

$$\text{conf}(Y' \Rightarrow (X \setminus Y')) < c.$$

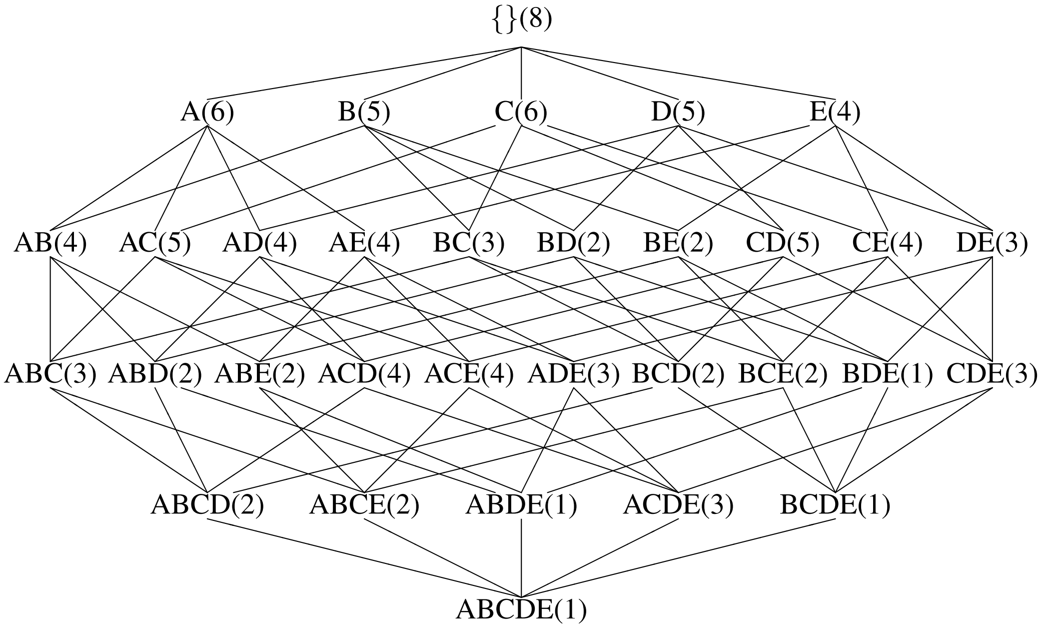
(a) Prove the theorem.

(b) Sketch an algorithm (pseudo code) that generates all association rules with support σ or above and a minimum confidence of c, provided the set F of all frequent itemsets (w.r.t. σ) with their support, efficiently using the pruning power of the given theorem.

Exercise 3-2 Support based on closed frequent itemsets (1 point)

(a) The database from the lecture grew by one transaction. We computed the corresponding support of all itemsets in the lattice:

TID	A	B	C	D	E
1	0	1	0	0	0
2	1	0	1	1	1
3	1	1	1	0	1
4	0	0	1	1	0
5	1	1	1	1	1
6	1	0	1	1	1
7	1	1	0	0	0
8	1	1	1	1	0



Identify the closed frequent itemsets for the support thresholds $\sigma = 4$ and $\sigma = 2$, respectively.

What do you observe?

(b) Sketch an algorithm (pseudo code) to find the support for all frequent itemsets, using only the set of closed frequent itemsets as information.

Exercise 3-3 Apriori (1 point)

Consider the following transaction database D over the items $I = \{A, B, C, D, E, F\}$.

TransID	Items
1	A B E
2	B D
3	C D F
4	A B D
5	A C E
6	B C E F
7	A C E
8	A B C E
9	A B C D F
10	B C D E

Given the support threshold $\sigma = 2$, apply the Apriori algorithm and extract all frequent itemsets w.r.t. the given threshold. Please explain in the solution all the steps that you followed.

In particular include for each level the candidate set (C_k) (i) after the join step before pruning and (ii) after pruning. Annotate for those objects pruned in (ii) the explicit reason for pruning them.

Also give explicitly the solution of frequent k-itemsets (S_k) for each k.