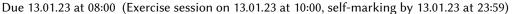
Exercise sheet #1: Frequent Itemset Mining



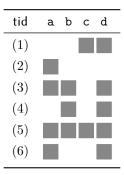


Please carefully read and follow the instructions regarding course work submissions. Failing to meet the requirements might lead to penalties. https://elearn.uef.fi/mod/page/view.php?id=135658

If you suspect that something is wrong with some exercise question, please contact the lecturer.

If you face persistent issues while working on an exercise, do ask for help, e.g. during a course meeting or by contacting the lecturer via email.

Consider the following dataset containing four items and six transactions, represented as a binary matrix.



Problem 1 (Dataset).

- a) Write down this dataset in the standard horizontal format, i.e. as a collection of item ids for each transaction.
- b) Write down this dataset in vertical format, i.e. as a collection of transaction ids for each item.

Problem 2 (Itemset lattice).

- a) Draw the itemset lattice, indicating the absolute support of each itemset.
- b) Mark closed itemsets and maximal frequent itemsets at absolute minimum support of $\sigma = 2$.

Problem 3 (Association rules).

- a) Determine the support set, as well as the absolute and relative support of itemsets $\{a\}$, $\{a,b\}$, $\{b,d\}$ and $\{a,b,d\}$.
- b) Determine the confidence of association rules $\{a,b\} \Rightarrow \{d\}, \{b,d\} \Rightarrow \{a\} \text{ and } \{a\} \Rightarrow \{b,d\}.$

Problem 4 (Apriori algorithm).

a) Run the level-wise apriori algorithm to mine all frequent itemsets at absolute minimum support of $\sigma=2$, assume an ordering of items by frequency. Show intermediate steps, i.e. all candidates generated and pruned, where relevant, as well as the corresponding enumeration tree.