

DM868/DM870/DS804: Data Mining and Machine Learning
Spring term 2023

Exercise 2: Apriori, Confidence, Itemsets and Association Rules

Exercise 2-1 Combinatoric explosion (1 point)

- (a) A database contains transactions over the following items: “apples”, “bananas”, and “cherries”. How many different combinations of these items can exist (i.e., how many different transactions could possibly occur in the database)?

(We do not distinguish whether a transaction contains a fruit once or several times, e.g., if someone bought one apple or several apples would just result in the transaction to contain “apples”.)
- (b) The database now also contains the items “dates”, “eggplants”, “figs”, and “guavas”. How many possible transactions do we have now?
- (c) How many combinations (possible different transactions) do we have with n items?
- (d) How many transactions with exactly two items (i.e., 2-itemsets) can we have when the database contains 3 items? When it contains 5 items? How many k -itemsets do we have when the database contains n items?

Exercise 2-2 Itemsets and Association Rules (1 point)

Given a set of transactions T according to the following table:

Set of transactions T	
Transaction ID	items in basket
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies }
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

- (a) What are the support and the confidence of $\{\text{Milk}\} \Rightarrow \{\text{Diapers}\}$?
- (b) What are the support and the confidence of $\{\text{Diapers}\} \Rightarrow \{\text{Milk}\}$?
- (c) What is the maximum number of size-3 itemsets that can be derived from this data set?
- (d) What is the maximum number of association rules that can be extracted from this dataset (including rules, that have zero support)?
- (e) What is the maximum size of frequent itemsets that can be extracted (assuming $\sigma > 0$)?
- (f) Find an itemset (of size 2 or larger) that has the largest support.
- (g) Find a pair of items, a and b , such that the rules $\{a\} \Rightarrow \{b\}$ and $\{b\} \Rightarrow \{a\}$ have the same confidence.

Exercise 2-3 Apriori candidate generation (1 point)

Given the frequent 3-itemsets:

$\{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 3, 4\}, \{1, 3, 5\}, \{2, 3, 4\}, \{2, 3, 5\}, \{3, 4, 5\}$

List all candidate 4-itemsets following the Apriori joining and pruning procedure.