



Karl M. Heuer / David E. Roberson

Autumn 2023
Exercise class:
in 302 on September 26th

EXERCISES FOR COMPUTATIONAL TOOLS FOR DATA SCIENCE (02807)

WEEK 5: FREQUENT ITEMSETS

Exercises from MMDS

6.1.1, 6.1.3, 6.1.5, 6.1.6, (6.1.7)
6.2.5, 6.2.6

Exercise 2: Implement the A-Priori algorithm

Implement a version of the A-Priori algorithm on your own. You may assume your data is given as a list of baskets.

Exercise 3: Use built in tools

Use/import the following Python packages: *Pandas* and *MLxtend*. Especially, have a look at *apriori* and *association_rules* from *mlxtend.frequent_patterns*. For documentation see:

<http://rasbt.github.io/mlxtend/>

If helpful / desirable you might also use *TransactionEncoder* from *mlxtend.preprocessing* to clean / prepare your data.

Work with the following data, which is provided as a csv-file (after uncompressing). It was published by Heeral Dedhia on kaggle.com in 2020 under a General Public License, version 2.: [groceries.zip](#)

The task: determine:

1. the frequent pairs of items.
2. the association rules of high confidence with or w/o high lift.
3. (optional) the association rules of high confidence with or w/o high interest. (optional)

Note: it seems the interest measure is not already implemented. You would have to do it yourself.

Experiment with values for *support*, *lift* and *interest*. E.g. vary *support* between 0,0002 and 0,02. Try a *lift* measure around 1,5 and then vary.