Lab 3 – Pattern & Association Mining #1

Data Mining, Spring 2017

Mathias Löwe (malw@itu.dk) | Daniel Sørensen (dafr@itu.dk)

IT UNIVERSITY OF COPENHAGEN

Today's Lab: Patterns!

Pattern & Association Mining #1

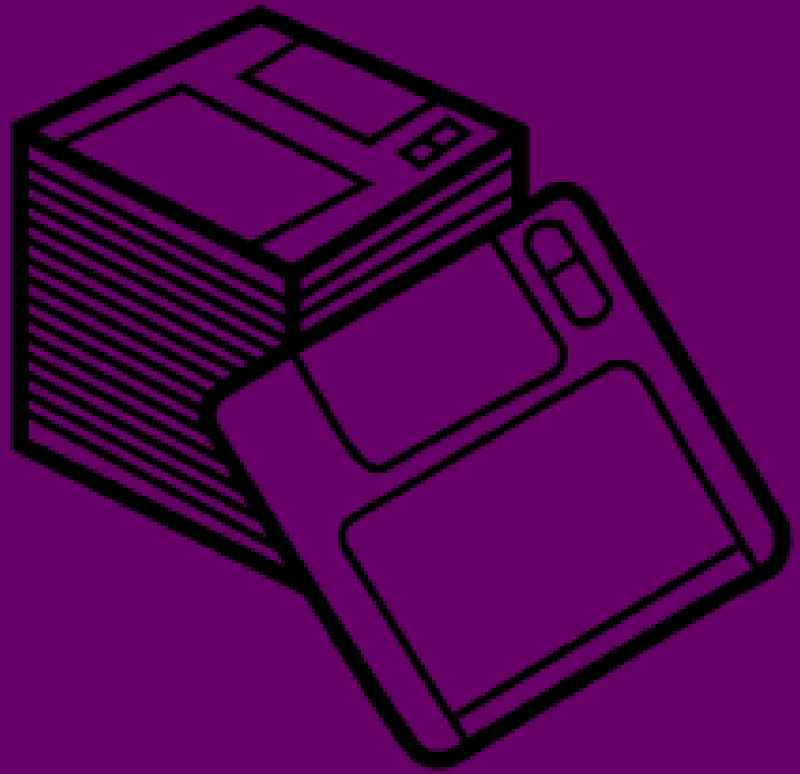
- Today you will be searching for frequent patterns in some simple transactional data.
- You will implement the apriori algorithm to accomplish this.
  - Page 248-254 (chapter 6.2.1-6.2.2) in the book.
- A simple code structure is provided to help you get started.

Code Provided

- Two classes
  - Apriori
  - ItemSet
- The ItemSet class is used to encapsulate information of sets of transaction items constructed during the algorithm.
- The Apriori class is where you should implement the algorithm.
  - Methods
    - Main
    - apriori
    - generateFrequentItemSets
    - joinSets
    - generateFrequentItemSetsLevel1
    - countSupport
- Code provided makes use of the HashTable java data structure, which is used to store <Key, Value> pairs. Values can then be retrieved based on their key. Is in this instance used to store <ItemSet, Integer> pairs, where the integer is used to store the support value for the item set.

The Data

- The transactional data is simple and is only made up of integers
- See it as different records of sales, where each number is an item with id=1, id=2 and so on.
- The data set is provided in the code as the TRANSACTIONS two-dimensional integer array in the Apriori class.
- Also included is the two-dimensional integer array BOOK_TRANSACTIONS in the Apriori class
  - Has the data the book uses in its example
  - Use this to check your implementation



Lab Overview

- First take a look at the code provided.
- Then start working on your apriori implementation
  - Suggested order of implementation of methods in the Apriori class:
    - countSupport
    - joinSets
    - generateFrequentItemSetsLevel1
    - generateFrequentItemSets
    - apriori
    - main

Thanks for listening!