Lab2 - Frequent Itemsets and Association Rules
Data Mining - KTH

Andrea Scotti, Daniele Montesi

15 November 2019

1 Introduction

The goal of the assignment is to find the frequent itemsets with support at least s, where s is given, and to generate association rules with confidence at least c, where c is given, from the itemsets found in the first step.

To test and evaluate the implementation, we wrote a program that uses a Apriori algorithm implementation to discover frequent itemsets with support at least s in a given dataset of sales transactions.

2 Code Explanation

The code is organized in a python notebook file. For every part of the lab, we created one or more functions. You can find each of them in a different cell of the python notebook. The the lab tasks are here explained.

1. find_k-itemsets(prev_itemsets, k): it constructs a sets of k-tuples, those that might be frequent sets (support > s) based on information from the pass for k–1.
2. filter_frequent_itemsets(itemsets, s): it constructs the set of truly frequent k-tuples, by filtering only those k-tuples from the given itemsets that have support at least s.
3. generate_association_rules(frequent_itemset, c): given a frequent itemset, it generates all the possible association rules with confidence at least c.

3 How to Run the Code

To run the code follow the steps above:

Requirements: Jupyter notebook, Python 3.6

1. open the terminal and run

```
> jupyter notebook
```

2. open the file "AssociationRules.ipynb" with the notebook
3. run all the cells

You should obtain the following output:

```
In [11]: list(C4.keys())

Out[11]: [(240, 274, 448, 538),
          (240, 274, 448, 834),
          (240, 274, 538, 834),
          (240, 448, 538, 834),
          (274, 448, 538, 834),
          (120, 205, 581, 834),
          (708, 853, 883, 978),
          (175, 597, 766, 910),
          (175, 597, 766, 960),
          (175, 597, 910, 960),
          (175, 766, 910, 960),
          (571, 623, 795, 853),
          (597, 766, 910, 960),
          (392, 461, 569, 862),
          (350, 411, 572, 579),
          (350, 411, 572, 803),
```

Figure 1: Itemsets found of length 4, choosing support s

```
In [47]: i = 1
         for rule in generate_association_rules([C1, C2, C3, C4], 0.8):
             print("Rule " + str(i) +" : ", rule[0].get_first(), "->", rule[0].get_second(), "\nConfidence:
         ", np.round(rule[1], 3))
             i+=1

         Rule 1 :  (653, 722, 887) -> 995
         Confidence: 0.89
         Rule 2 :  (205, 285, 461) -> 829
         Confidence: 0.949
         Rule 3 :  (32, 93, 350) -> 937
         Confidence: 0.95
         Rule 4 :  (205, 285, 461) -> 529
         Confidence: 0.949
         Rule 5 :  (32, 93, 350) -> 606
         Confidence: 0.935
         Rule 6 :  (205, 285, 403) -> 950
         Confidence: 0.979
         Rule 7 :  (32, 72, 350) -> 937
         Confidence: 0.932
         Rule 8 :  (205, 285, 403) -> 829
         Confidence: 0.946
         Rule 9 :  (32, 72, 350) -> 606
```

Figure 2: First associations rules found choosing confidence 0.8