HW2: Discovery of Frequent Itemsets and Association Rules

Authors: Andrei Capastru, Alex Porciani

The task is to implement algorithms that: find frequent itemsets given a support(lower threshold) and then create association rules given a specific confidence.

1.1 How to run the code

0. Install python3 and install Jupyter Notebook
1. Run in unix terminal:

```
cd ~
git clone https://github.com/andreicap/data-mining.git
cd data-mining/HW2
jupyter notebook
```

2. Open the browser on `http://localhost:8888` and run the `frequent_itemsets.ipynb` file

1.2 Finding frequent itemsets

In this phase we create itemsets and count their frequency in the dataset. The ratio between frequency and the total number of rows in the dataset represent the support of an itemset.

1. Generate itemsets of length k=1.
2. Calculate the support for each of them and retain only the ones with support above the set threshold
3. Generate itemsets of lenght k+1
4. Calculate the support for each of them and retain only the ones with support above the set threshold
5. Repeat step 3 and 4 until there are no itemsets selected above the threshold.

1.3 Association rules

In this phase we calculate the confidence for all the association rules and keep only those above the set confidence.

1. Generate all the association rules from the frequent itemset.
2. Calculate confidence for all the association rules.
3. Retain those who are above set confidence threshold.