CS F415 - Data Mining
Semester 2 - 2019-2020

Sheet 2
Topic: Association Rule Mining

1 Research Papers

Some interesting papers for rule mining have been listed below. Two of them are the original papers in which Apriori and FP-Growth algorithms were proposed. The third one is a recent take on frequent itemset mining in the form of a review of the developments in the field in the past 25 years.

1. Apriori: http://www.rsrikant.com/papers/vldb94_rj.pdf
   Citation: Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." In Proc. 20th int. conf. very large data bases, VLDB, vol. 1215, pp. 487-499. 1994.

2. FP Growth: https://dl.acm.org/doi/pdf/10.1145/335191.335372
   Citation: Han, Jiawei, Jian Pei, and Yiwen Yin. "Mining frequent patterns without candidate generation." ACM sigmod record 29, no. 2 (2000): 1-12.

3. Frequent Itemset Mining Review:
   http://philippe-fournier-viger.com/25yearsFIM.pdf

2 Links to help you with code

1. Apriori:

   - A detailed tutorial (along with code) on how to run association analysis for a large dataset using Python generators:

https://www.kaggle.com/datatheque/association-rules-mining-market-basket-analysis

- A notebook with exploratory data analysis and apriori on Black Friday shopping data:
https://www.kaggle.com/dabate/black-friday-examined-eda-apriori

- Github repository containing easy-to-follow simple code from scratch:
https://github.com/asaini/Apriori/

2. FP Growth:

- A Github repo with complete code for implementing FP growth from scratch:
https://github.com/calee0219/Python3-Fp-growth