Lab Course Machine Learning

Exercise Sheet 9

Prof. Dr. Dr. Lars Schmidt-Thieme
Jung Min Choi

HiWi: Harish Malik
Submission deadline : December 19, 2024

1 Market Basket Analysis (5 points)

1. **[1 Point]** Read the `basket.csv` and preprocess the dataset for further analysis.
2. **[4 Points]** Compute and report:
   - Frequency of each itemset.
   - Support of each itemset.
   - Confidence of association rules.
   - Lift of association rules.

   Note: External libraries like `sklearn` or prebuilt packages for association rule mining are not allowed.
   Note: Only association rules between one-item sets (e.g., `item1` $\rightarrow$ `item2`) should be considered for this task.

2 Apriori (10 points)

1. Download the dataset from the *link*. Examine the dataset and perform necessary preprocessing.
2. **[5 Points]** Implement the vanilla Apriori algorithm from scratch. Print the frequent itemsets for each support level and the corresponding association rules.
3. **[5 Points]** Implement the improved Apriori algorithm from scratch. You can refer the improved Apriori algorithm from the Machine Learning 1 Lecture Slides. Print the frequent itemsets for each support level and the corresponding association rules.

Note: Use of external libraries (e.g., `sklearn` or prebuilt Apriori packages) is not allowed.
Note: Set the minimum support level to 0.02 and the minimum confidence level to 0.3.

3 Eclat (5 points)

1. **[5 Points]** Implement the vanilla Eclat algorithm from scratch. Print the frequent itemsets for each support level and the corresponding association rules.

Note: Use of external libraries (e.g., `sklearn` or prebuilt Eclat packages) is not allowed.
Note: Set the minimum support level to 0.02 and the minimum confidence level to 0.3.