Association Rule Mining using R

CIS400/600 Fundamentals of Data and Knowledge Mining

Note: Provide necessary R commands for questions that start with *. For other questions, include the answers in the same R file with comments

Problem 1 - Data! Data! Data!

Use the classic **Groceries** dataset to perform association analysis. The dataset is provided as **CSV** file. The dataset has been provided along with the exercise as **groceries.csv**

1. * Install **'arules'** package and import it

2. * Process the dataset and summarize it

3. How many transactions have 1 item in them

4. What is the maximum length of a transaction

Problem 2 - Association analysis

1. * Print the item with minimum item frequency

2. * Print the item with maximum item frequency

3. * Print first 10 items with minimum item frequency

4. * Plot the first 10 items with maximum item frequency

Problem 3 - Association analysis (Apriori)

Parameters: support = 0.01, confidence = 0.05, minimum length = 2
Answer the questions below with respect to these parameters

1. * What is the R command to generate rules using Apriori algorith with above specified parameters

2. How many rules are generated

3. * Display the generated rules

4. * Sort the rules according to support and print the first 10 rules with highest support

5. * Sort the rules according to lift and print the first 10 rules with highest lift

6. * Among the generated rules, list the rules that have "whole milk" on the right-hand side

7. * Among the generated rules, list the rules that have "whole milk" on the left-hand side

8. * Among the generated rules, list the rules that only have "whole milk" and "soda" on the left-hand side

9. How many rules are generated in part 8

10. How many rules are generated with support = 0.5