Christopher Singh Data Mining Assignment 5

Apriori Rule Analysis

Dataset: <https://www.kaggle.com/roshansharma/market-basket-optimization>

Data Pre-processing – No data cleaning was needed since all the columns and rows had the appropriate values.

Steps Taken For Apriori Implementation in R:

* First install the arules and arulesViz packages.
* Then load the libraries into the workspace.
* Read the csv file using the basket format and a comma as a separator.
* Now use the built-in Apriori function and pass the loaded dataframe along with the appropriate parameters.
* The list of parameters that I used were support, confidence, target and maxlength. I ran three different Apriori algorithms with different support and confidence values. I used maxlength to limit the number of items in the frequency.
* The last step is to use the inspect method on the result to see the rules.

Results when Support was 1% and Confidence was 20%:

* Since the lift parameter for most of the rules were greater than 1, there is a positive relationship between the items.
* The lift was never negative for any of the rules.
* The top grocery products that had the most rules were: mineral water, cooking oil and chicken.

Results when Support was 5% and Confidence was 25%:

* There were only 4 rules found. The items were chocolate, mineral water, eggs, and spaghetti.
* The right-hand side was mostly mineral water.
* There was also a positive relationship between all the rules.

There were no results when the support was 50% and the confidence was 50%.

My interpretation of the results:

* Mineral water seemed to be the most common grocery product purchased as it appeared the most in my first 2 attempts of Apriori.
* I never found a negative relationship between the products because the lift was always greater than 0.
* Whenever I try to make the confidence level too high, I always seem to end up with an empty list as a result.