THEORY

Apriori algorithm is used for finding frequent itemsets in a dataset for association rule mining. It is called Apriori because it uses prior knowledge of frequent itemset properties. We apply an iterative approach or level-wise search where k-frequent itemsets are used to find k+1 itemsets. To improve the efficiency of the level-wise generation of frequent itemsets an important property is used called Apriori property which helps by reducing the search space.

* **Support:** Support is an indication of how frequently the itemset appears in the dataset. It is the count of records containing an item ‘x’ divided by the total number of records in the database.
* **Confidence:** Confidence is a measure of times such that if an item ‘x’ is bought, then item ‘y’ is also bought together. It is the support count of (x U y) divided by the support count of ‘x’.
* **Lift:** Lift is the ratio of the observed support to that which is expected if ‘x’ and ‘y’ were independent. It is the support count of (x U y) divided by the product of individual support counts of ‘x’ and ‘y’.

### **Algorithm**

1. Read each item in the transaction.
2. Calculate the support of every item.
3. If support is less than minimum support, discard the item. Else, insert it into frequent itemset.
4. Calculate confidence for each non- empty subset.
5. If confidence is less than minimum confidence, discard the subset. Else, it into strong rules.

R is a language that is developed to support statistical calculations and graphical computing/ visualizations. It has an in-built library function called **arules** which implements the Apriori algorithm

‘**arules**’ package provides the infrastructure for representing, manipulating, and analyzing transaction data and patterns.

’**arulesviz**’ package is used for visualizing Association Rules and Frequent Itemsets.

‘**apriori()**‘ function is in-built in R to mine frequent itemsets and association rules using the Apriori algorithm. Here, ‘Income’ is the transaction data. ‘parameter’ is a named list that specifies the minimum support and confidence for finding the association rules

**inspect()** function prints the internal representation of an R object or the result of an expression