Apriori Algorithm

An Unsupervised Machine Learning Algorithm

Written by – Urdhva Borse

What is it?

The Apriori algorithm is an unsupervised machine learning algorithm used for associate rule learning on databases – especially those containing transactions. What is *associate rule learning?* It is a rule-based machine learning method for discovering interesting relations between variables in large databases. In any given transaction with a variety of items, association rules are meant to discover the rules that determine how or why certain items are connected.

How does it work? (Oversimplified version)

1. Generate Itemset

An itemset is a group of items with their frequencies. The algorithm uses unique items and combines them based on a probability of occurrence above a certain threshold. The algorithm sorts out itemsets that are not as frequent (to work on meaningful itemsets only).

2. Expand and then prune itemsets

Using the aptly named *Apriori property* – that states that if an itemset appears frequently in a dataset, all its subsets must also appear frequently – the algorithm combines frequent itemsets further to form larger itemsets.

3. Repeat step 1 & 2

The algorithm repeats steps 1 & 2 until all frequent itemsets meeting the defined threshold probability are generated exhaustively. Each iteration generates more complex and comprehensive associations in the itemsets.

Once Apriori has created the itemsets, the strength of the generated associations and relationships can be investigated.

Where do you use this algorithm?

The most common use - data mining to help market research on frequently purchased items in all kinds of shops.

Healthcare – It is used to identify a combination of drugs and patient factors that are associated with adverse drug reactions.