Association rules

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occurs in a transaction. A typical example is a Market Based Analysis.

Market Based Analysis is one of the key techniques used by large relations to show associations between items.

It allows retailers to identify relationships between the items that people buy together frequently.

For example, if a customer buys bread, he most likely can also buy butter, eggs, or milk, so these products are stored within a shelf or mostly nearby.

Association rule learning can be divided into three types of algorithms:

1. **Apriori**
2. **Eclat**
3. **F-P Growth Algorithm**

How does Association Rule Learning work?

Association rule learning works on the concept of If and Else Statement, such as if A then B.

Here the If element is called **antecedent**, and then statement is called as **Consequent**. These types of relationships where we can find out some association or relation between two items is known *as single cardinality*. It is all about creating rules, and if the number of items increases, then cardinality also increases accordingly. So, to measure the associations between thousands of data items, there are several metrics. These metrics are given below:

- **Support**
- **Confidence**
- **Lift**

Support

Support is the frequency of A or how frequently an item appears in the dataset. It is defined as the fraction of the transaction T that contains the itemset X. If there are X datasets, then for transactions T, it can be written as:

$$\text{Supp}(X) = \frac{Freq(X)}{T}$$

Confidence

Confidence indicates how often the rule has been found to be true. Or how often the items X and Y occur together in the dataset when the occurrence of X is already given. It is the ratio of the transaction that contains X and Y to the number of records that contain X.

$$\text{Confidence}=\frac{Freq(X,Y)}{Freq(X)}$$

Lift

It is the strength of any rule, which can be defined as below formula:

$$\text{Lift}=\frac{Supp(X,Y)}{Supp(X)\times Supp(Y)}$$

It is the ratio of the observed support measure and expected support if X and Y are independent of each other.

Algorithm Association rule generation from a frequent itemset.

INPUT Z A frequent itemset
INPUT *min_conf* the minimum confidence threshold
for all item i in Z **do**
    Construct a rule $Z - \{i\} \Rightarrow \{i\}$
    if *confidence*$(Z - \{i\} \Rightarrow \{i\}) \geq min\_conf$ **then**
        output $Z - \{i\} \Rightarrow \{i\}$
        add $\{i\}$ to the set C_1
Set $k = 2$
repeat
    for all itemset V of size k generated by joining two itemsets from C_{k-1} *do*
        Construct a rule $Z - V \Rightarrow V$
        if *confidence*$(Z - V \Rightarrow V) \geq min\_conf$ **then**
            output $Z - V \Rightarrow V$
            add V to the set C_k
until $k < Z - 1$

Applications of Association Rule Learning

It has various applications in machine learning and data mining. Below are some popular applications of association rule learning:

- **Market Basket Analysis:** It is one of the popular examples and applications of association rule mining. This technique is commonly used by big retailers to determine the association between items.
- **Medical Diagnosis:** With the help of association rules, patients can be cured easily, as it helps in identifying the probability of illness for a particular disease.
- **Protein Sequence:** The association rules help in determining the synthesis of artificial Proteins.
- It is also used for the **Catalog Design** and **Loss-leader Analysis** and many more other applications.

Cross support patterns

Cross-support patterns occur when you have patterns that show up that relate high frequency items with low frequency items and their association is due to the high frequency item.
a ratio smaller than a set threshold.

Normally many found patterns are cross-support patterns which contain frequent as well as rare items.