Everything on Association Rules

TID	*Items*
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

Itemset : A collection of one or more items

E.g {Bread, Milk, Beer}

K-itemset : An itemset that contains k items

Support Count (σ) : Frequency of occurrence of an itemset

E.g σ({Bread, Milk, Diaper}) = 2

Support (s) : Fractions of transactions containing an itemset

E.g s({Bread, Milk, Diaper}) = 2/5

Frequent Itemset : An itemset whose support is greater than or equal to a minimum support threshold

Association Rules are expressed in the form:

$X \rightarrow Y$, where X and Y are itemsets

The support (s) of a transaction is the % of transactions that contain both X and Y
E.g $s = \frac{\sigma(Milk, Diaper, Beer)}{|T|}$ = ⅖ = 0.4

The confidence © of a role is the percentage of transactions that contain X, that also contain Y
E.g $c = \frac{\sigma(Milk, Diaper, Beer)}{\sigma(Milk, Diaper)}$ = ⅔ = 0.67

The direction of the transaction is important

X→Y != Y→X

Rules originating from the same itemset have same support but different confidence

General Equations :

$$s = \frac{\sigma(x+y)}{|T|}$$

$$c = \frac{\sigma(x+y)}{\sigma(x)}$$

Discard all the rules that have a confidence score lower than some predefined target