Basket Analysis

Market basket analysis or affinity is process of identifying and extracting meaningful pattern from transaction sets.

Support

The support of rule is a fraction of transaction that contain the rule. Support is to determine which rule are potentially actionable, we need to first identify which rule are significant or interesting and limit our focus to those alone.

$$\textbf{Support}_{\{bread,cereal\}\rightarrow\{milk\}} = \textbf{Support}_{\{milk\}\rightarrow\{bread,cereal\}}$$

Support

Customer	Items Bought
A	fruit, bread, cereal
B	fruit, milk, cereal, eggs
C	bread, milk, cereal, cheese
D	fruit, bread, milk, cereal
E	fruit, bread, milk, cheese

- Support { bread,cereal} –{milk} =2/5=0.4
- 40% of all transactions include bread, cereal and milk

confidence

The confidence of a rule is the Predictive power or accuracy of the rule

$$\text{Confidence}_{X \to Y} = \frac{\text{Support}_{X \to Y}}{\text{Support}_{X}}$$

Confidence

TID	Items Bought
A	fruit, bread, cereal
B	fruit, milk, cereal, eggs
C	bread, milk, cereal, cheese
D	fruit, bread, milk, cereal
E	fruit, bread, milk, cheese

- Confidence $_{\{\text{bread, cereal}\}-\{\text{milk}\}}$ =0.4/0.6=0.67
- Off all the transaction where both bread and cereal were purchased, 67% of them also included the purchase of milk

Lift

the lift of a rule is the increase likelihood of the rule occurring relative to its typical rate of occurrence

$$\text{Lift}_{X \to Y} = \frac{\text{Confidence}_{X \to Y}}{\text{Support}_{Y}}$$

Lift

TID	Items Bought
A	fruit, bread, cereal
B	fruit, milk, cereal, eggs
C	bread, milk, cereal, cheese
D	fruit, bread, milk, cereal
E	fruit, bread, milk, cheese

- Lift $_{\{\text{bread, cereal}\} \rightarrow \{\text{milk}\}}$ =0.67/0.8=0.84
- Customer who brought bread and cereal were 16%(1-0.84%) less likely to also bye milk

Leverage

Leverage is difference between the observed frequency of a rule and frequency that would be expected if X and Y were independent.

$$\text{Leverage}_{X \to Y} = \text{Support}_{X \to Y} - \text{Support}_X \times \text{Support}_Y$$

Leverage

TID	Items Bought
A	fruit, bread, cereal
B	fruit, milk, cereal, eggs
C	bread, milk, cereal, cheese
D	fruit, bread, milk, cereal
E	fruit, bread, milk, cheese

- Leverage $_{\{\text{bread, cereal}\} \rightarrow \{\text{milk}\}}$ =0.4/-0.6 * 0.8=-0.08
- Customer who brought bread and cereal were slightly less likely to also buy milk

Conviction

Conviction quantifies how dependent the consequent is on the antecedent.

$$\text{Leverage}_{X \to Y} = \text{Support}_{X \to Y} - \text{Support}_X \times \text{Support}_Y$$

Conviction

TID	Items Bought
A	fruit, bread, cereal
B	fruit, milk, cereal, eggs
C	bread, milk, cereal, cheese
D	fruit, bread, milk, cereal
E	fruit, bread, milk, cheese

- Conviction $_{\{\text{bread, cereal}\} \rightarrow \{\text{milk}\}} = 1\text{-}0.8/1\text{-}0.67 = -0.61$
- The rule would be correct 39%(1-0.61) more often if the consequent were dependent on the antecedent