Association Rule

- What It Does?
  - Determine which behaviors/outcomes go together
  - Find relationships among attributes in data that frequently occur together

- Examples
  - Market basket analysis: determine what things go together in a shopping cart at the supermarket (we will do this in the later part of the lecture today)
  - Trajectory: where do tourists go after arriving at las vegas? Where go tourists go after going to a casino (maybe a nearby hotel)
  - Disease or symptom development overtime

Market-Basket Analysis

- What It Does?
  - Data-mining technique for determining sales patterns
  - Shows products that customers tend to buy together

- Examples
  - On Thursday nights, people who buy diapers may also buy beer

No one predicted that result. After seeing the results of the data mining, Wal-Mart moved the beer next to the diapers and beer sales went up.

- Mark Whitehorn 2006

Market-Basket Analysis

- A total of N transaction records

- Support Count: Number of times that (transactions in which) certain product(s) has(have) been bought. For example:
  - n(A) = # transactions in which A was bought
  - n(A&B) = # of transactions in which both A and B were bought

- Support: Probability that a certain product (or a group of products) is (are) bought during a transaction. For example:
  - For example: P(A), P(A&B)

$$\text{Support(A)} = P(A) = \frac{n(A)}{N} \qquad \text{Support(A\&B)} = P(A\&B) = \frac{n(A\&B)}{N}$$

Market-Basket Analysis

- Confidence: given that a person is buying product A, the likelihood he/she will also buy product B (in the same transaction)
  - n(A) = # transactions in which A was bought
  - n(A&B) = # of transactions in which both A and B were bought

$$P(B|A) = \frac{P(A\&B)}{P(A)} = \frac{\frac{n(A\&B)}{N}}{\frac{n(A)}{N}} = \frac{n(A\&B)}{n(A)}$$

- Lift: the ratio of confidence to the base probability (support) of buying an item
  - The "lift" in probably of buying B if one is buying A too.

$$\frac{P(B|A)}{P(B)} = \frac{\textit{Probability of buying B given that they're buying A}}{\textit{Probability of buying B anyway}}$$

Market-Basket Analysis

Tran #	Milk	Diapers	Beer
1	1	0	0
2	0	1	1
3	1	1	1
4	1	1	1
5	1	1	0

- Association rule $\underbrace{\{\text{Milk, Diapers}\}}_{\text{Antecedent}} \rightarrow \underbrace{\{\text{Beer}\}}_{\text{Consequent}}$

Market-Basket Analysis

- Confidence

$$P(\text{Consequent}|\text{Antecedent}) = \frac{P(\text{Conseqent\&Antecedent})}{P(\text{Antecedent})}$$

- Lift:

$$\text{Lift} = \frac{P(\text{Consequent}|\text{Antecedent})}{P(\text{Consequent})}$$

Market-Basket Analysis

Tran #	Milk	Diapers	Beer
1	1	0	0
2	0	1	1
3	1	1	1
4	1	1	1
5	1	1	0

- Rule: {Milk, Diapers} → {Beer}

- Supports
  - n(Milk &Diapers) = 3 P(Milk & Diapers) = 3/5 = 0.6
  - n(Beer) = 3 P(Beer) = 3/5 =0.6
  - n(M&D&B) = 2 P(M&D&B) = 2/5 = 0.4

Market-Basket Analysis

Tran #	Milk	Diapers	Beer
1	1	0	0
2	0	1	1
3	1	1	1
4	1	1	1
5	1	1	0

- Rule: {Milk, Diapers} → {Beer}
- Confidence
  - n(M&D&B) /n(Milk &Diapers) = P(M&D&B) / P(Milk & Diapers) = 0.66667
- Lift
  - Confidence / P(Beer) = 0.66667 / 0.6 = 1.111111……
- What did we learn
  - Folks who are buying Milk and Diapers are more likely to buy beer (compared to the average customer)
- Sometimes we need thresholds to mine the best rules.

Algorithm Used to Find Association Rules

- Apriori (most common)
  - Step 1: calculate support for single-item itemsets
  - Step 2: only keep itemsets with supp > minsup
  - Step 3: expand to two item itemsets
  - Breadth-first search (BFS)
  - The problem is exponentially growing computation time

- Eclat (equivalence class transformations)
  - Step 1: calculate a single-item itemset
  - Step 2: if supp>minsup then add item
  - Step 3: if in some step supp<minsup, move to another single-item itemset
  - Deepth-first search (DFS)