ML – 9

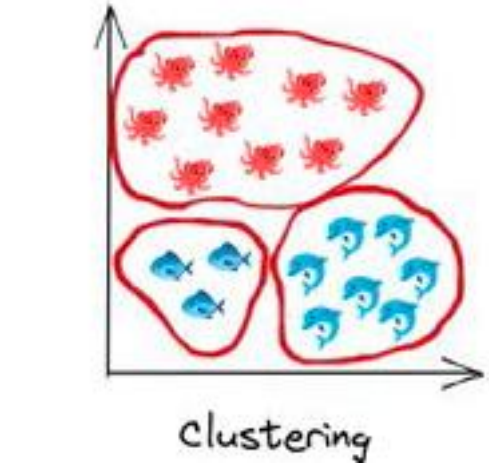
Machine Learning

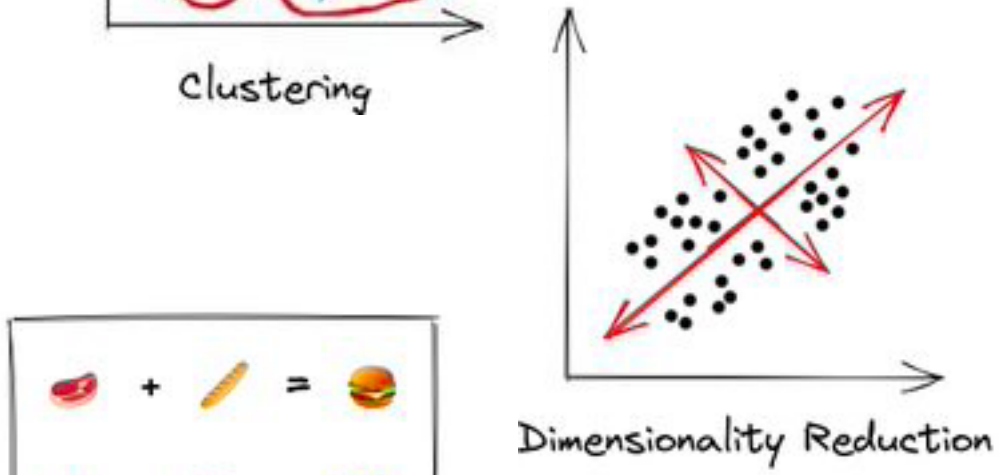
- Unsupervised Learning
  - Association Rules
    - Apriori Algorithm

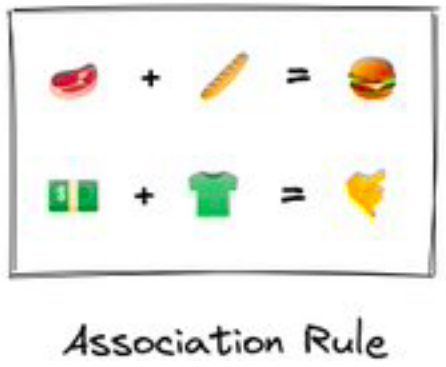
Prof. Umut ORHAN

Unsupervised Learning

1. Clustering
   + K-means
2. Dimensionality Reduction
   + PCA
3. **Association Rules**
   + Apriori



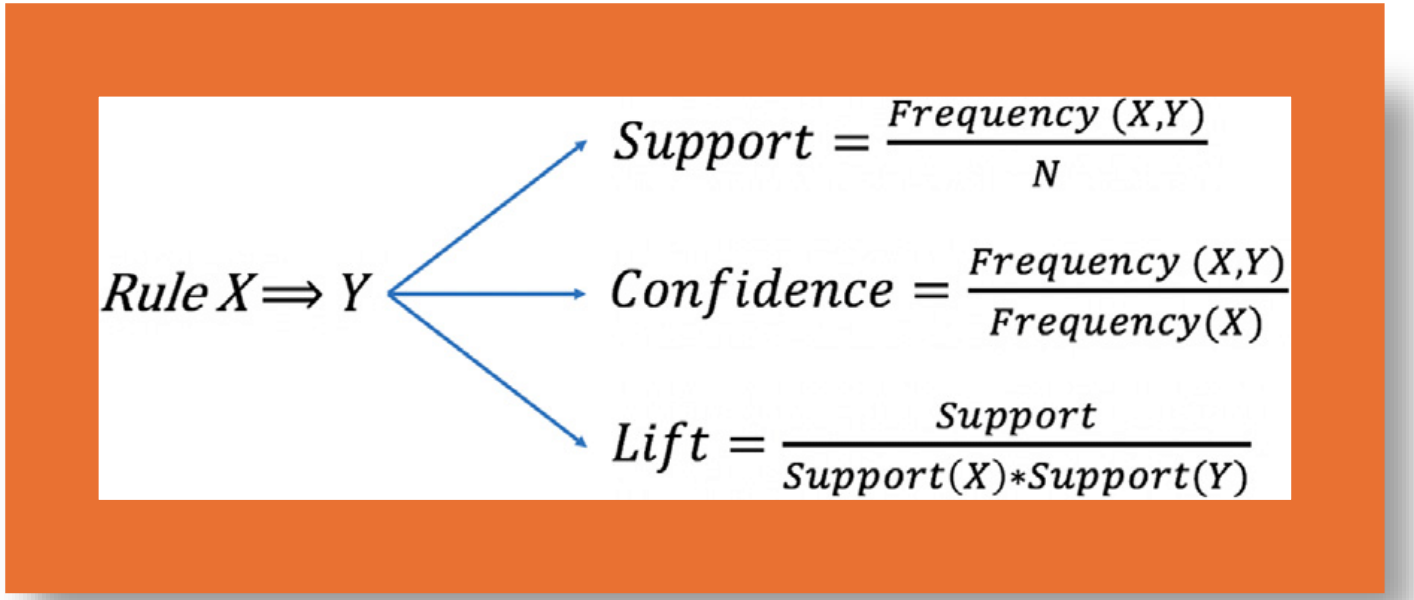




Apriori- Real World Applications

- Market Basket Analysis
  - Retail
  - E-Commerce
  - Supermarkets
- Healthcare
- Finance
- Web Usage Mining

Apriori - Terms



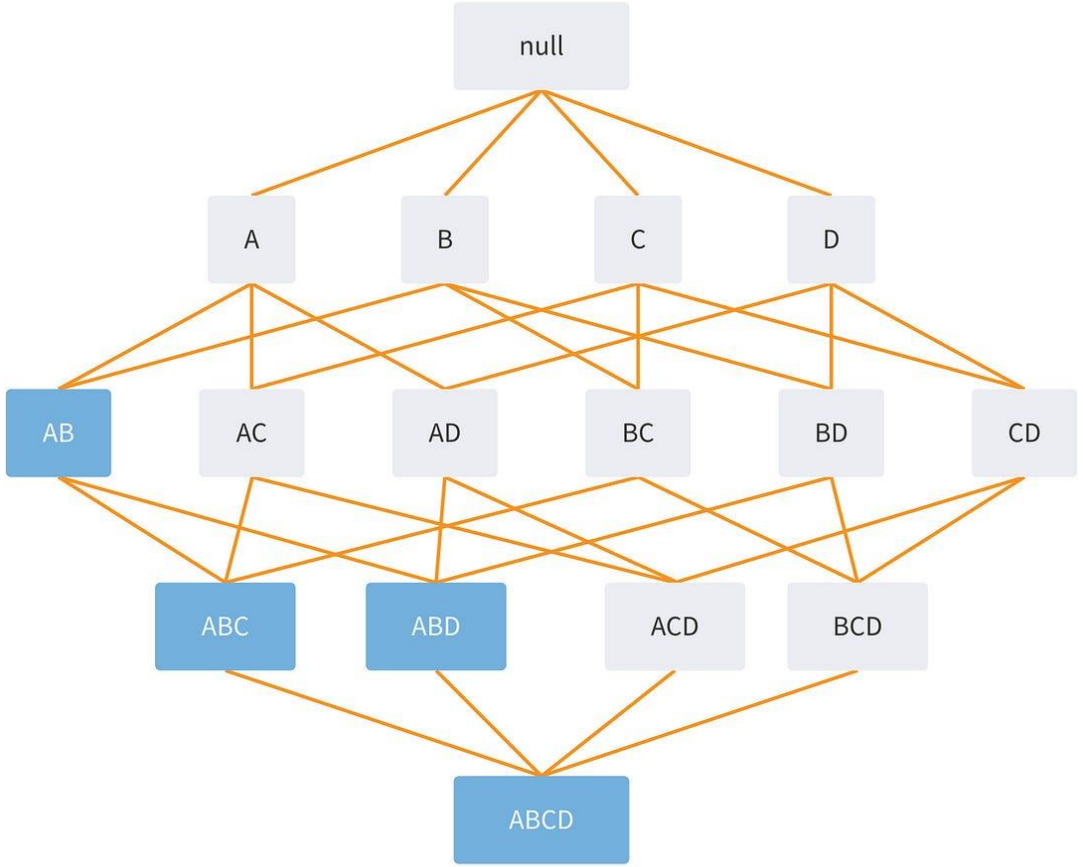
Support measures the frequency with which an itemset appears in the dataset.

Confidence represents the probability that the consequent of a rule occurs, given the antecedent.

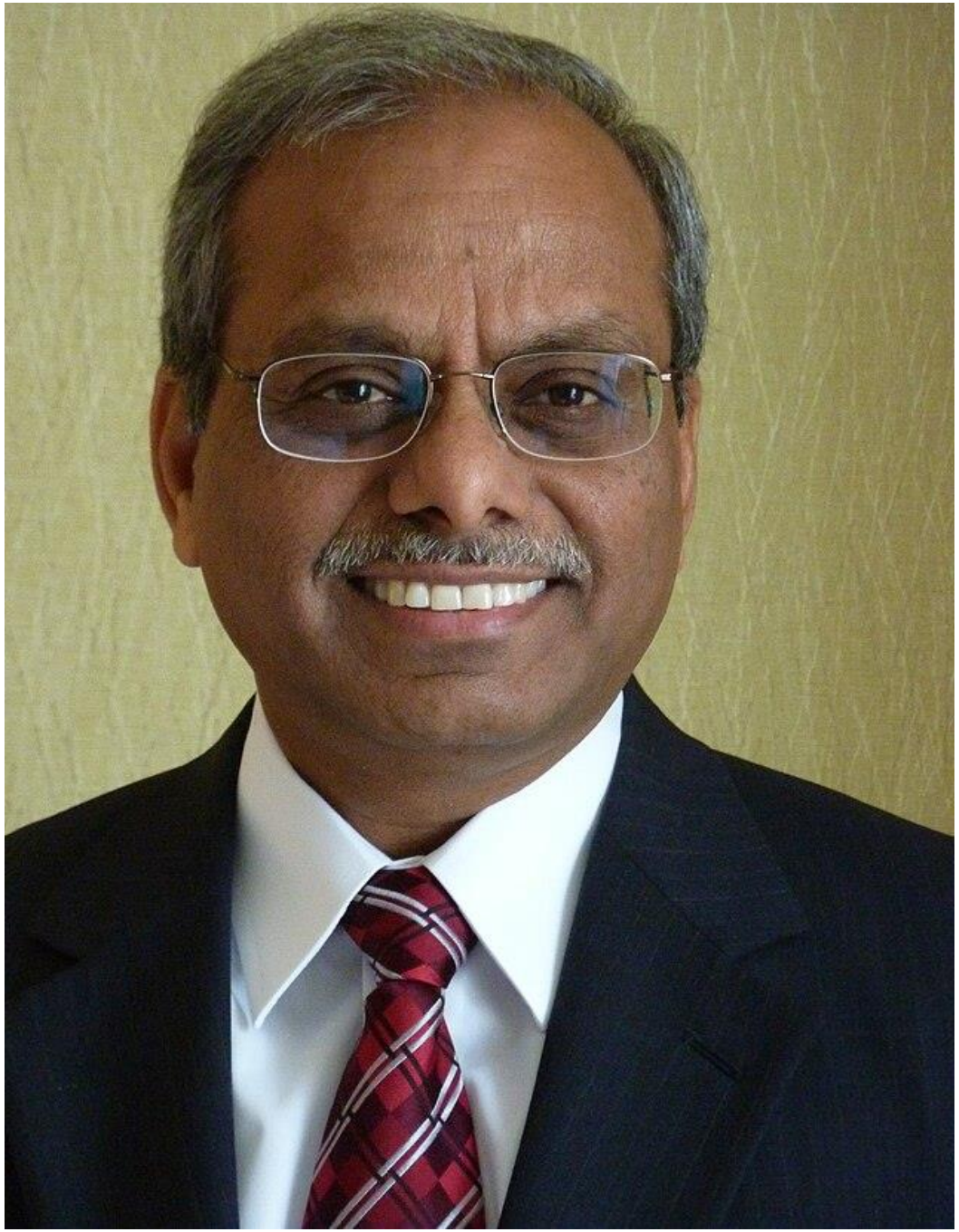
Lift evaluates how much more likely the consequent is to occur compared to random chance, given the antecedent.

Apriori Algorithm

1. Compute frequencies of items and filter them with a support threshold.
2. Join the items, compute their frequencies and filter again.
3. Repeat steps 1 and 2 until support value not available
4. Compute confidence and lift values.



Rakesh Agrawal



Apriori Advantages

1. Simplicity
2. Wide Applicability
3. Efficient for Small to Medium Datasets
4. Frequent Itemset Discovery

Apriori Disadvantages

1. Computational Cost
2. Low Efficiency
3. Threshold Selection
4. Repetitive Calculations
5. Ignoring Rare Items

T ID	Items Bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Beer, Diapers}

Determined that:

min_support = %40,
min_confidence = %70.

Step 1:

T ID	Items Bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Beer, Diapers}

1-Itemset	Support_count
Bread	3
Butter	3
Milk	2
Beer	2
Cookies	1
Diapers	3

Bread, Butter, Milk, Diapers, Beer

1-Frequent Itemset	Support_count
Bread	3
Butter	3
Diapers	3
Milk	2
Beer	2

Step 2:

T ID	Items Bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Beer, Diapers}

2-Itemset	Support_count
Bread, Butter	3 ✓
Bread, Diapers	1
Bread, Milk	2 ✓
Bread, Beer	0
Butter, Diapers	1
Butter, Milk	2 ✓
Butter, Beer	0
Diapers, Milk	1
Diapers, Beer	2 ✓
Milk, Beer	0

Bread, Butter, Diapers, Milk, Beer

2-Frequent Itemset	Support_count
Bread, Butter	3
Bread, Milk	2
Butter, Milk	2
Diapers, Beer	2

Step 3:

T ID	Items Bought
1	{Bread, Butter, Milk}
2	{Bread, Butter}
3	{Beer, Cookies, Diapers}
4	{Milk, Diapers, Bread, Butter}
5	{Beer, Diapers}

Bread, Butter, Milk, Diapers, Beer

3-Itemset	Support_count
Bread, Butter, Milk	2 ✓
Bread, Butter, Diapers	1
Bread, Butter, Beer	0
Bread, Milk, Diapers	1
Bread, Milk, Beer	0
Bread, Diapers, Beer	0
Butter, Milk, Diapers	1
Butter, Milk, Beer	0
Butter, Diapers, Beer	0
Milk, Diapers, Beer	0

3-Frequent Itemset	Support_count
Bread, Butter, Milk	2

Example of Apriori-1

min_confidence is 70%

• Confidence is:: **(X→Y) = P(Y | X) = P(X U Y) / P(X)**

• We have 5 frequent itemsets: → {Bread, Butter}, {Bread, Milk}, {Butter, Milk}, {Diapers, Beer} and {Bread, Butter, Milk}.

1-Frequent Itemset	Support_count
Bread	3
Butter	3
Diapers	3
Milk	2
Beer	2

2-Frequent Itemset	Support_count
Bread, Butter	3
Bread, Milk	2
Butter, Milk	2
Diapers, Beer	2

3-Frequent Itemset	Support_count
Bread, Butter, Milk	2

- For {Bread, Butter},
  - bread>butter = 3/3 = 100% (Strong) ✓
  - butter>bread = 3/3 = 100% (Strong) ✓

- For {Bread, Milk}
  - bread>milk = 2/3 = 67% ✗
  - milk>bread = 2/2 =100% (Strong) ✓

- For {Butter, Milk}
  - butter>milk = 2/3 = 67% ✗
  - milk>butter = 2/2 = 100% (Strong) ✓

- For {Diapers, Beer}
  - diapers>beer = 2/3 = 67% ✗
  - beer>diapers = 2/2 = 100% (Strong) ✓

- For {Bread, Butter, Milk}
  - bread,butter>milk = 2/3 = 67% ✗
  - bread,milk>butter = 2/2 = 100% (Strong) ✓
  - milk,butter>bread = 2/2 = 100% (Strong) ✓
  - bread>butter,milk = 2/3 = 67% ✗
  - butter>bread,milk = 2/3 = 67% ✗
  - milk>bread,butter = 2/2=100% (Strong) ✓

Example of Apriori-2

The Lift Formula

Example Association Rule: Bread > Milk

$$lift = \frac{P(X \cap Y)}{P(X) * P(Y)} \rightarrow Lift(\text{Bread} \rightarrow \text{Milk}) = \frac{P(\text{Bread} \cap \text{Milk})}{P(\text{Bread}) \cdot P(\text{Milk})}$$

$P(\text{Bread} \cap \text{Milk})$ = Support_count / Total Transaction → **2 / 5 = 0.4**

$P(\text{Bread})$ = Support_count / Total Transaction → **3 / 5 = 0.6**

$P(\text{Milk})$ = Support_count / Total Transaction → **2 / 5 = 0.4**

→ $\frac{0.4}{(0.4) * (0.6)} = 1.67$ → **1.67 > 1** → **Positive Correlation** ✓

Example of Apriori-3

An example of a lift value for an association rule

Machine Learning

9. week

Thanks for watching