Data and Data Sources

Market Basket analysis

- It identifies associations between products in transactions.
- Uses Association Rule Mining to generate rules like "If a customer buys X, they are likely to buy Y."
- Commonly applied in retail, e-commerce, and recommendation systems.

Association Analysis

Association Analysis is a data mining technique used to **discover relationships or patterns** between items in large datasets. It is widely used in **market basket analysis, recommendation systems, fraud detection, and web usage mining**.

Objective:

To find frequent item-sets and association rules that describe how items are related within a dataset.

Association Rules

Association rules are statements in the form of:

If $X \Rightarrow Y$, Which means, **if item X appears, item Y is also likely to appear**.

Association Analysis

Example:

- **{Bread, Butter}** → **{Milk}** (People who buy bread and butter often buy milk)
- **{Laptop}** → **{Mouse}** (People who buy a laptop are likely to buy a mouse)

Key metrics used to evaluate association rules:

1. Support

Measures how frequently an itemset appears in the dataset.

$$Support(X) = \frac{\text{Frequency of X in dataset}}{\text{Total transactions}}$$

Association Analysis

Example: If **Milk** appears in **30 out of 100 transactions**, then:

$$Support(Milk) = \frac{30}{100} = 30\%$$

2. Confidence

Measures how often **Y** appears when **X** is present.

$$Confidence(X \Rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

Example: If **Bread** appears in **50 transactions**, and in **40** of them, **Milk** is also bought:

$$Confidence(Bread \Rightarrow Milk) = \frac{40}{50} = 80\%$$

UPES
UNIVERSITY OF TOMORROW

Association Analysis

3. Lift

Measures how much **stronger** the association is compared to a random occurrence.

$$Lift(X \Rightarrow Y) = \frac{\text{Confidence}(X \rightarrow Y)}{\text{Support}(Y)}$$

If Lift > 1: X and Y are positively correlated (buying one increases the likelihood of buying the other).

If Lift < 1: X and Y are negatively correlated (buying one reduces the likelihood of buying the other).

Problems

A retailer wants to analyze buying patterns based on **500 transactions** in a week:

- **{Laptop} appears in 100 transactions.**
- **{Laptop, Mouse} together appear in 60 transactions.**
- **{Mouse} appears in 150 transactions.**

Questions:

1. What is the **confidence** of the rule **{Laptop}** → **{Mouse}**?
2. What is the **confidence** of the rule **{Mouse}** → **{Laptop}**?

UPES
UNIVERSITY OF TOMORROW

Problems: Supermarket Transactions

Transaction Dataset

Transaction ID	Items Purchased
T1	Milk, Bread, Butter
T2	Bread, Butter
T3	Milk, Bread
T4	Milk, Bread, Butter, Eggs
T5	Bread, Butter, Eggs

Step 1: Compute Support

- Support(Milk)
- Support(Bread)
- Support(Butter)
- Support({Milk, Bread})
- Support({Bread, Butter})

Step 2: Compute Confidence

- Confidence(Milk → Bread)
- Confidence(Bread → Butter)

Step 3: Compute Lift

- Lift(Milk → Bread)
- Lift(Bread → Butter)

Problems

Transaction Data

Transaction ID	Items Purchased
T1	Apple, Banana, Milk
T2	Apple, Banana
T3	Apple, Banana, Milk
T4	Banana, Milk, Bread
T5	Apple, Bread
T6	Banana, Bread
T7	Apple, Banana, Bread

Lift(Apple → Banana) 0.87

Lift(Banana → Bread) 0.60

UPES
UNIVERSITY OF TOMORROW

Applications of Market Basket analysis

Retail:

- Optimize product placement (**e.g., placing Milk near Bread**).
- Identify frequently bought-together items for promotions.

E-commerce & Recommendations:

- Suggest items frequently bought together (**Amazon's "Customers who bought this also bought..."**).
- Improve personalized recommendations.

Healthcare: Analyze patient symptoms and medications that are frequently prescribed together.

Finance: Detect fraud by identifying unusual spending patterns.

Practice

Q. Using the following transactional dataset of customer purchases. Find:

i. Frequent Itemset/s

ii. Association rules

iii. Support, confidence and lift of the rules

Transaction ID	Items Purchased
1	Bread, Milk, Eggs
2	Bread, Butter
3	Milk, Butter
4	Bread, Milk, Butter, Cheese
5	Eggs, Milk
6	Bread, Eggs
7	Milk
8	Bread, Butter, Milk

Practice

Frequent Itemsets (Let's use a minimum support of 2):

Transaction ID	Items Purchased
1	Bread, Milk, Eggs
2	Bread, Butter
3	Milk, Butter
4	Bread, Milk, Butter, Cheese
5	Eggs, Milk
6	Bread, Eggs
7	Milk
8	Bread, Butter, Milk

- **Individual Items:**
  - Bread: 5/8 = 0.625 (Support = 0.625)
  - Milk: 6/8 = 0.75 (Support = 0.75)
  - Eggs: 3/8 = 0.375 (Support = 0.375)
  - Butter: 4/8 = 0.5 (Support = 0.5)
  - Cheese: 1/8 = 0.125 (Support = 0.125)
- **Pairs:**
  - {Bread, Milk}: 3/8 = 0.375 (Support = 0.375)
  - {Bread, Butter}: 3/8 = 0.375 (Support = 0.375)
  - {Milk, Butter}: 3/8 = 0.375 (Support = 0.375)
  - {Milk, Eggs}: 2/8 = 0.25 (Support = 0.25)
  - {Bread, Eggs}: 2/8 = 0.25 (Support = 0.25)
- **Triplets:**
  - {Bread, Milk, Butter}: 2/8 = 0.25 (Support = 0.25)

Practice

Support, Confidence, and Lift:

- **Support:** The proportion of transactions that contain the itemset.
- **Confidence:** The probability that a transaction containing A also contains B (A -> B).
- **Lift:** The ratio of the observed support to the support if A and B were independent. A lift greater than 1 suggests a positive association.

Association Rules (Using the frequent itemsets):

- **{Bread, Milk} -> {Butter}:**
  - Support = 2/8 = 0.25
  - Confidence = 2/3 = 0.666
  - Lift = (2/8) / ((3/8) * (4/8)) = 1.33
- **{Bread, Butter} -> {Milk}:**
  - Support = 2/8 = 0.25
  - Confidence = 2/3 = 0.666
  - Lift = (2/8) / ((3/8) * (6/8)) = 0.888
- **{Milk, Butter} -> {Bread}:**
  - Support = 2/8 = 0.25
  - Confidence = 2/3 = 0.666
  - Lift = (2/8) / ((3/8) * (5/8)) = 1.066

Transaction ID	Items Purchased
1	Bread, Milk, Eggs
2	Bread, Butter
3	Milk, Butter
4	Bread, Milk, Butter, Cheese
5	Eggs, Milk
6	Bread, Eggs
7	Milk
8	Bread, Butter, Milk

Practice

Transaction ID	Items Purchased
1	Bread, Milk, Eggs
2	Bread, Butter
3	Milk, Butter
4	Bread, Milk, Butter, Cheese
5	Eggs, Milk
6	Bread, Eggs
7	Milk
8	Bread, Butter, Milk

Rule	Support	Confidence	Lift
{Bread} -> {Milk}	0.375	0.6	0.8
{Bread} -> {Butter}	0.375	0.6	1.2
{Milk} -> {Bread}	0.375	0.5	0.8
{Milk} -> {Butter}	0.375	0.5	1
{Butter} -> {Bread}	0.375	0.75	1.5
{Butter} -> {Milk}	0.375	0.75	1.25
{Bread, Milk} -> {Butter}	0.25	0.666	1.33
{Bread, Butter} -> {Milk}	0.25	0.666	0.888
{Milk, Butter} -> {Bread}	0.25	0.666	1.066