7.09 APRIORI

APRIORI Association Rule Models

- It is a popular ML method that deploys pattern recognition to identify and quantify relationships between different, yet related items

- It is also known as an **association rule mining** modelling method

APRIORI Association Rule Models

- Commonly used for market basket analysis and bundle recommendations
- Data sources include transaction data or clickstream data
- Key Levers: Support, Confidence & Lift

APRIORI Model Constraints

- Support
  - Refers to the default popularity of an item
  - Support(B) = (Transactions containing (B))/(Total Transactions)
- Confidence
  - Likelihood that Item B is also bought if item A is bought
  - Confidence(A→B) = (Transactions containing both (A and B))/(Transactions containing A)
- Lift
  - Likelihood Item B is purchased when item A is purchased, while controlling for how popular item B is.
  - Lift(A→B) = (Confidence (A→B))/(Support (B))

APRIORI Model - Support

Measure 1: Support. This says how popular an itemset is, as measured by the proportion of transactions in which an itemset appears. In Table 1 below, the support of {apple} is 4 out of 8, or 50%. Itemsets can also contain multiple items. For instance, the support of {apple, beer, rice} is 2 out of 8, or 25%.

$$\text{Support } \{🍎\} = \frac{4}{8}$$

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	💉 🍺 🍚 🍗
Transaction 6	💉 🍺 🍚
Transaction 7	💉 🍺
Transaction 8	💉 🍐

Table 1. Example Transactions

APRIORI Model – Confidence & Lift

Measure 2: Confidence. This says how likely item Y is purchased when item X is purchased, expressed as {X -> Y}. This is measured by the proportion of transactions with item X, in which item Y also appears. In Table 1, the confidence of {apple -> beer} is 3 out of 4, or 75%.

$$\text{Confidence}\{🍎 \rightarrow 🍺\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍎\}}$$

One drawback of the confidence measure is that it might misrepresent the importance of an association. This is because it only accounts for how popular apples are, but not beers. If beers are also very popular in general, there will be a higher chance that a transaction containing apples will also contain beers, thus inflating the confidence measure. To account for the base popularity of both constituent items, we use a third measure called lift.

Measure 3: Lift. This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is. In Table 1, the lift of {apple -> beer} is 1,which implies no association between items. A lift value greater than 1 means that item Y is *likely* to be bought if item X is bought, while a value less than 1 means that item Y is *unlikely* to be bought if item X is bought.

$$\text{Lift}\{🍎 \rightarrow 🍺\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍎\} \times \text{Support}\{🍺\}}$$

APRIORI Use Case

- Example: How likely is it that a customer that purchases eggs, will purchase bread too?

- Popular Use Case: product placement optimisation at both brick & mortar and e-commerce stores

APRIORI Use Case

- Example: eggs and bread are frequently purchased together. With this finding, you can increase sales by:
  - Placing eggs and bread next to each other so that when a customer buys one of the products, they don’t have to walk to buy the other product
  - Advertising to buyers of either eggs or butter in order to increase that person’s propensity to purchase the (bundled) other product
  - Offer discounts on both eggs and butter if the customer buys both of them in one purchase

APRIORI Association Rule Expression

- Association Rule:
  - “If eggs are purchased, the probability of buying bread is _____”
- Can also be represented as:
  - {eggs} → {bread}

Advantages of Association Rules

- Fast
- Works with small data
- Minimal feature engineering required

Measuring Association

- Support
- Confidence
- Lift

Measuring Association - Example

- Scenario: 5,000 total transactions in a supermarket
  - A = Bread purchases = 500 transactions
  - C = Egg purchases = 350 transactions
  - (A → C) Both bread and eggs purchased = 150 transactions

Measuring Association – Example (Continued)

Support

- support(bread) = (transactions containing bread)/(total number of transactions)
- support(bread) = 500 / 5000 = 0.1

Confidence

- confidence(bread → eggs) = (150/5000) / (500/5000) = 0.3 = 30%
- There is a 30% likelihood that eggs will be purchased if bread is purchased

Lift

- lift(bread → eggs) = 0.3 / (350/5000) = 4.28
- It is reasonable to state that a customer is likely to buy eggs if he has bought bread

Measuring Association – Example (Continued)

- Lift score > 1:
  - A is highly associated with C.
  - If A is purchased, it is likely that C will also be purchased
- Lift score < 1:
  - If A is purchased, it is unlikely that C will be purchased
- Lift score = 1:
  - Indicates that there is no association between items A and C