



# Introduction to Association Rules

In this module, we describe the unsupervised learning method called Association Rules. This is also called as affinity analysis or market-basket analysis.

## Association Rules

- Study of “what goes with what”
  - “Customers who bought X also bought Y”
  - What symptoms go with what diagnosis
- Transaction-based or event-based
- Also called “market basket analysis” and “affinity analysis”
- Originated with study of customer transactions databases to determine associations among items purchased

In association rules, the goal is to identify item clusters in transaction-type databases. Association rule discovery in marketing is termed “market basket analysis” and is aimed at discovering which groups of products tend to be purchased together. These items can then be displayed together, offered in post-transaction coupons, or recommended in online shopping.

## Terminology

- **Market basket transactions:**

t1: {bread, cheese, milk}

t2: {apple, eggs, salt, yogurt}

...

...

tn: {biscuit, eggs, milk}

- **Concepts:**

**An item:** an item/article in a basket

**I:** the set of all items sold in the store

**A transaction:** items purchased in a basket; it may have TID (transaction ID)

**A transactional dataset:** A set of transactions

Here, we summarize the basic terminology. A customer purchases items. A set of items is called a transaction. The set of transactions is a transactional dataset.

## Example 1

**The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) 2nd Edition**

by Trevor Hastie (Author), Robert Tibshirani (Author), Jerome Friedman (Author)

★★★★☆ 162 customer reviews

1st book in the **in biostatistics** series

**eTextbook** from \$49.35

**Hardcover** \$40.79 - \$45.37

**Paperback** \$112.05

**Other Sellers** See all 7 versions

**Buy used** \$40.79

Only 7 left in stock - order soon.

Ships from and sold by Galaxy\_Books.

**Note:** Not eligible for Amazon Prime. Available with free Prime shipping from other sellers on Amazon.

Get it Fri, May 10 - Fri, May 17

Get it Wed, May 8 if you choose paid shipping at checkout.

Deliver to: Houston, TX 77058

\$45.37 + Free Shipping

**Add to Cart**

**Buy Now**

Customers who bought this item also bought

An Introduction to Statistical Learning with Applications in R...  
★★★★☆ 207  
Hardcover  
\$40.00

Pattern Recognition and Machine Learning...  
★★★★☆ 168  
Hardcover  
\$45.15

Applied Predictive Modeling...  
★★★★☆ 85  
Hardcover  
\$55.25

Machine Learning & Probabilistic Reasoning...  
★★★★☆ 56  
Hardcover  
\$57.80

Deep Learning: Deep Neural Networks...  
★★★★☆ 109  
Hardcover  
\$37.95

Hands-On Machine Learning with Scikit-Learn and TensorFlow...  
★★★★☆ 109  
Paperback  
\$29.49

Reinforcement Learning: An Introduction...  
★★★★☆ 15  
Hardcover  
\$55.00

Page 1 of 15

WWW.KENT.EDU

Association rules, or affinity analysis, constitute a study of “what goes with what.” This method is also called market basket analysis because it originated with the study of customer transactions databases to determine dependencies between purchases of different items. Association rules are heavily used in retail for learning about items that are purchased together, but they are also useful in other fields. For example, a medical researcher might want to learn what symptoms appear together.

Association rules are commonly encountered in online recommendation systems (or recommender systems), where customers examining an item or items for possible purchase are shown other items that are often purchased in conjunction with the first item(s). The display from Amazon.com’s online shopping system illustrates the application of rules like this under “Frequently bought together.”

## Example 2

Transaction	Faceplate Colors Purchased				
1	red	white	green		
2	white	orange			
3	white	blue			
4	red	white	orange		
5	red	blue			
6	white	blue			
7	white	orange			
8	red	white	blue	green	
9	red	white	blue		
10	yellow				



Let us now consider a sample transactional dataset consisting of 10 transactions. The objective here is to discover if certain faceplate colors are consistently purchased together.

## Many Rules are Possible

For example: Transaction 1 supports several rules, such as

- “If red, then white” (“If a red faceplate is purchased, then so is a white one”)
- “If white, then red”
- “If red and white, then green”
- + several more

Managers often want to know if certain items are purchased together. These influence stocking and promotion decisions. Association rules provide information of this type in the form of “if–then” statements. These rules are computed from the data; unlike the if–then rules of logic, association rules are probabilistic in nature.

## Frequent Item Sets

- Ideally, we want to create all possible combinations of items
- **Problem 1:** computation time grows exponentially as # items increases.
- **Problem 2:** not all combinations may have business value.
- **Solution:** consider only “frequent item sets”

The idea behind association rules is to examine all possible rules between items in an if-then format, and select only those that are most likely to be indicators of true dependence. We use the term antecedent to describe the IF part, and consequent to describe the THEN part. In association analysis, the antecedent and consequent are sets of items (called itemsets) that are disjoint (do not have any items in common). Note that itemsets are not records of what people buy; they are simply possible combinations of items, including single items.

The first step in association rules is to generate all the rules that would be candidates for indicating associations between items. Ideally, we might want to look at all possible combinations of items in a database with  $p$  distinct items (in the phone faceplate example,  $p = 6$ , the number of colors of faceplates). This means finding all combinations of single items, pairs of items, triplets of items, and so on, in the transactions database. However, generating all these combinations requires a long computation time that grows exponentially in  $p$ . A practical solution is to consider only combinations that occur with higher frequency in the database. These are called frequent itemsets. Determining what qualifies as a frequent itemset is related to the concept of support, which we will study in future modules.

## Applications of Association Rule Mining

- Market Basket Analysis
- Medical Diagnosis
- Bioinformatics & Protein Sequence
- Fraud Detection

This concludes our introduction to Association Rule Mining.