

---

---

# Knowledge Discovery & Data Mining

## — Data Mining Tasks—

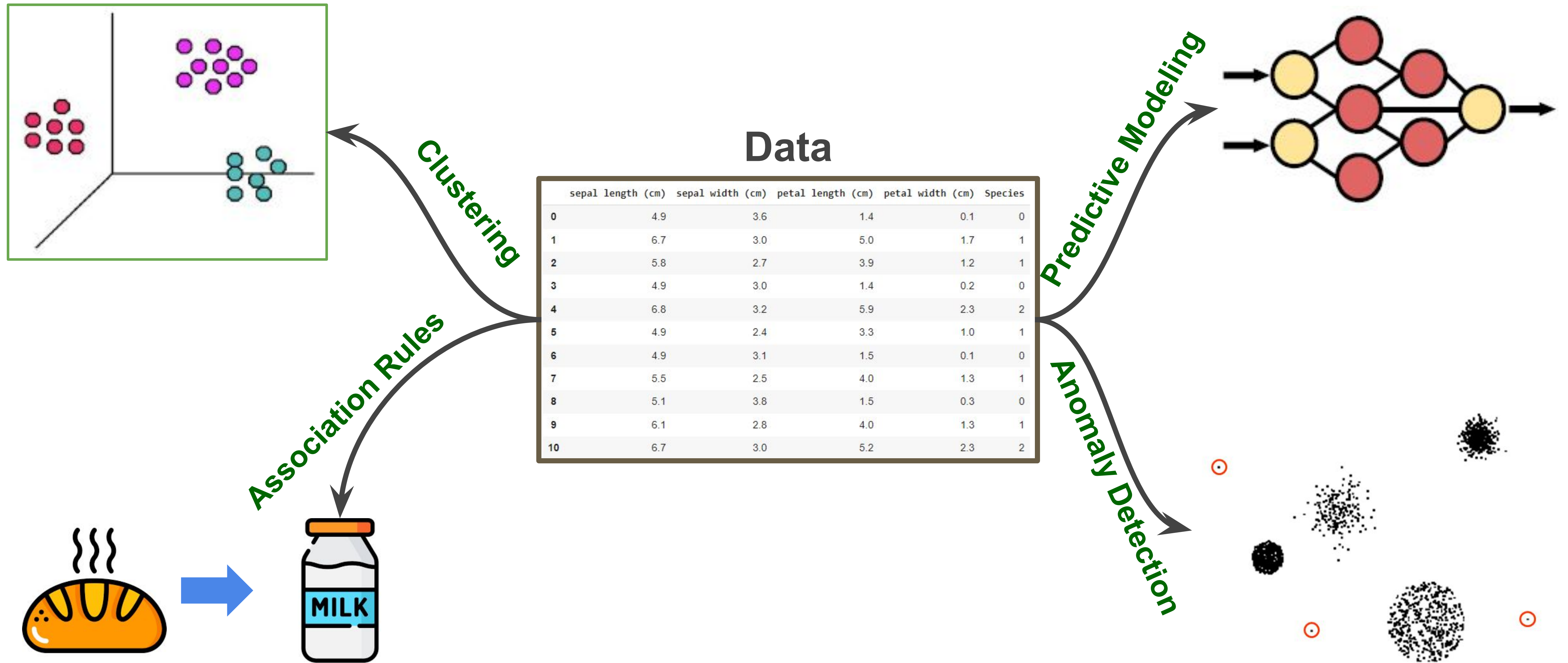
**Instructor: Yong Zhuang**

[yong.zhuang@gvsu.edu](mailto:yong.zhuang@gvsu.edu)

# Outline

- Common data mining tasks
  - Predictive modeling
    - Classification
    - Regression
    - Ranking
  - Clustering
  - Association rule mining
  - Anomaly detection

# Four common data mining tasks



# Predictive Modeling

**Goal:** Predict the value of an attribute based on the values of other attributes.

- ▶ The attribute to be predicted is often called the target attribute (also known as the dependent variable or response variable).
- ▶ The attributes used to make the prediction are often called explanatory attributes (also known as independent variables or predictors).

## Examples

- ▶ Predicting future price of a stock
- ▶ Predicting the annual rainfall at a location for the next 20 years
- ▶ Predicting whether a customer will buy something at a website
- ▶ Predicting who should be a friend of whom
- ▶ Predicting which web page to display when a user entered a search query



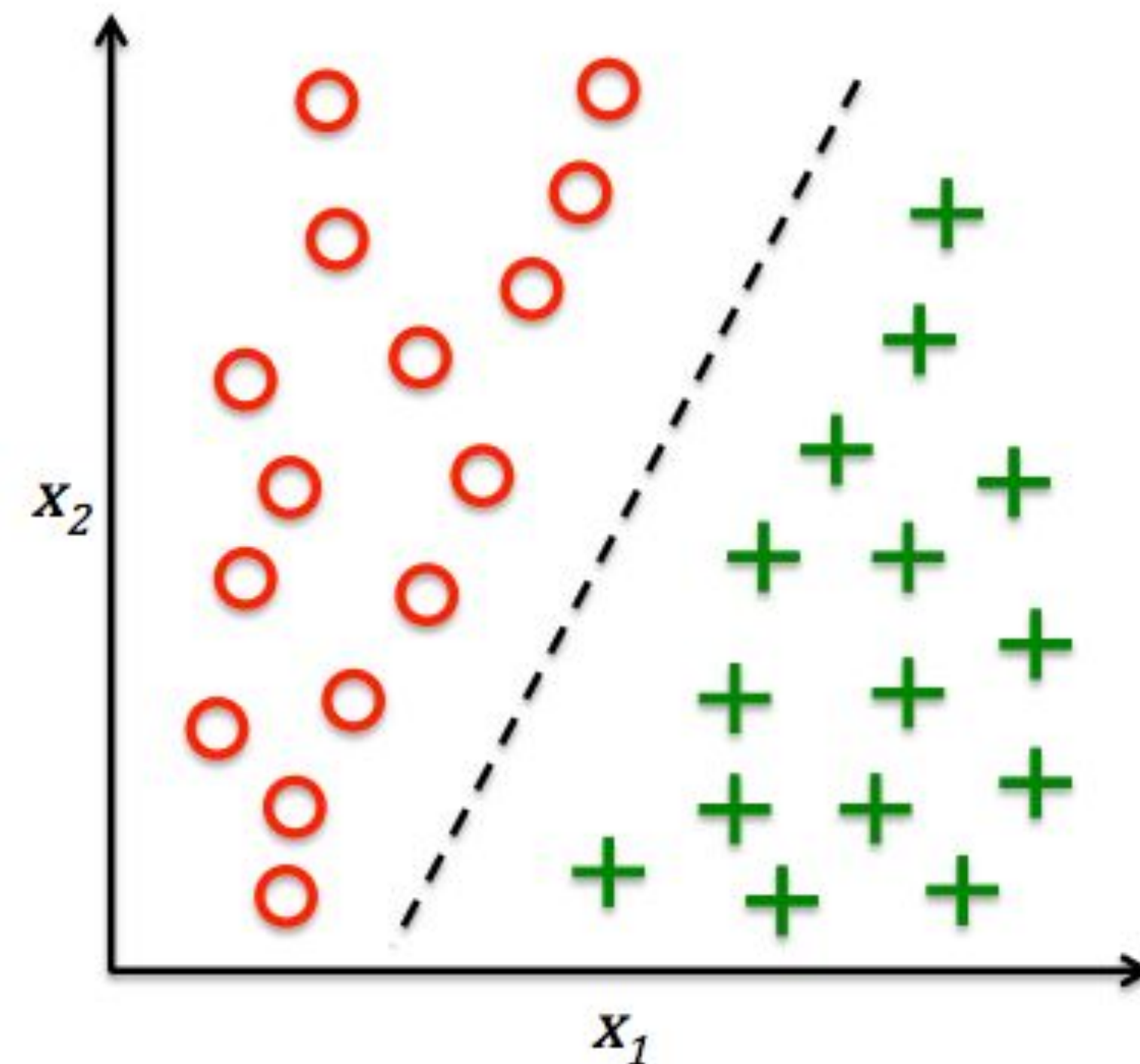
# Predictive Modeling: Classification

Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts.

- The model is derived based on the analysis of a set of training data (i.e., data objects for which the class labels are known).
- The model is used to predict the class labels of objects for which the class labels are unknown.

Examples:

- Text categorization
- Image classification
- Medical diagnosis
- Spam detection



# Predictive Modeling: Classification

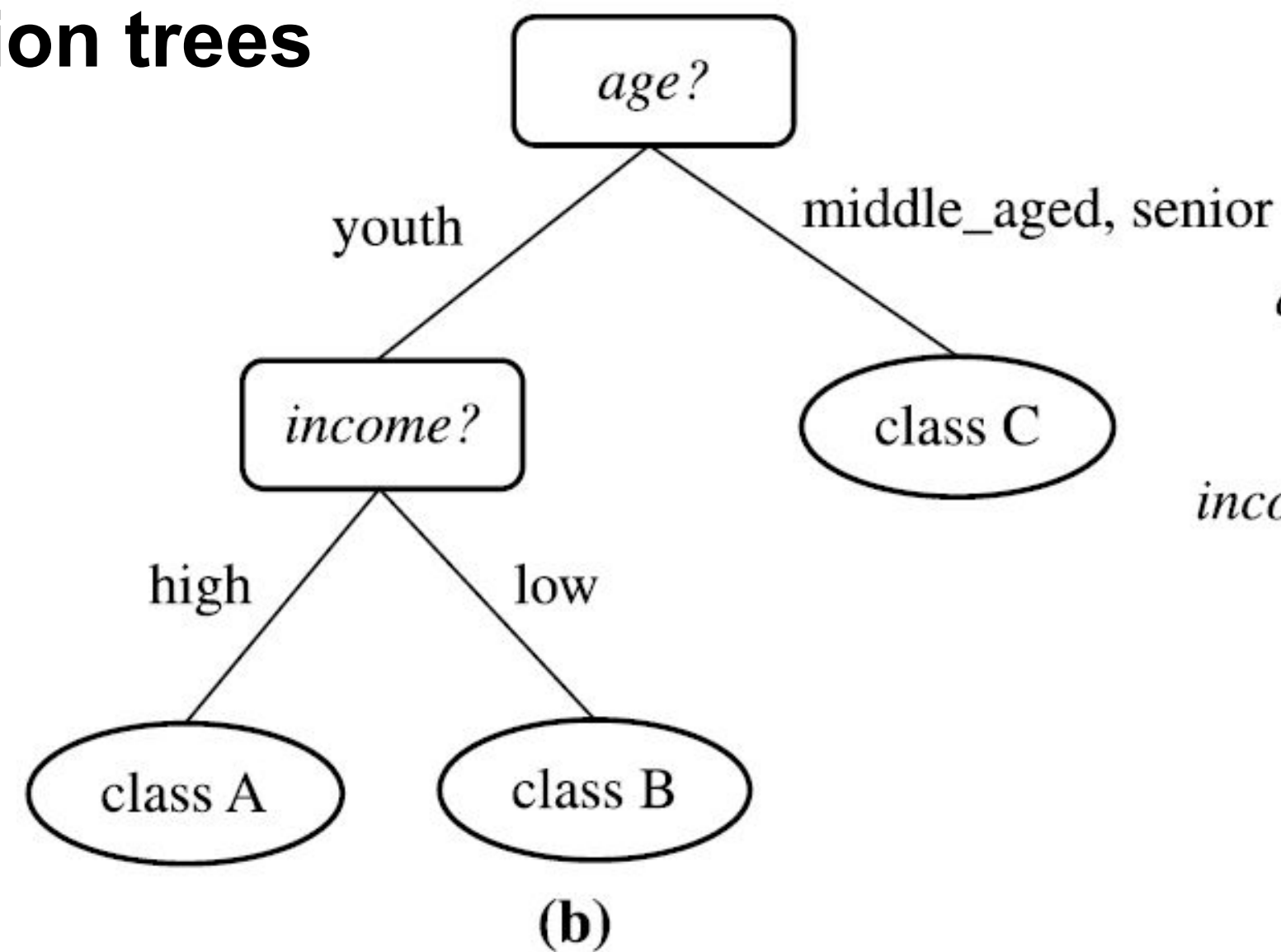
Typical methods:

## Rule-based classification

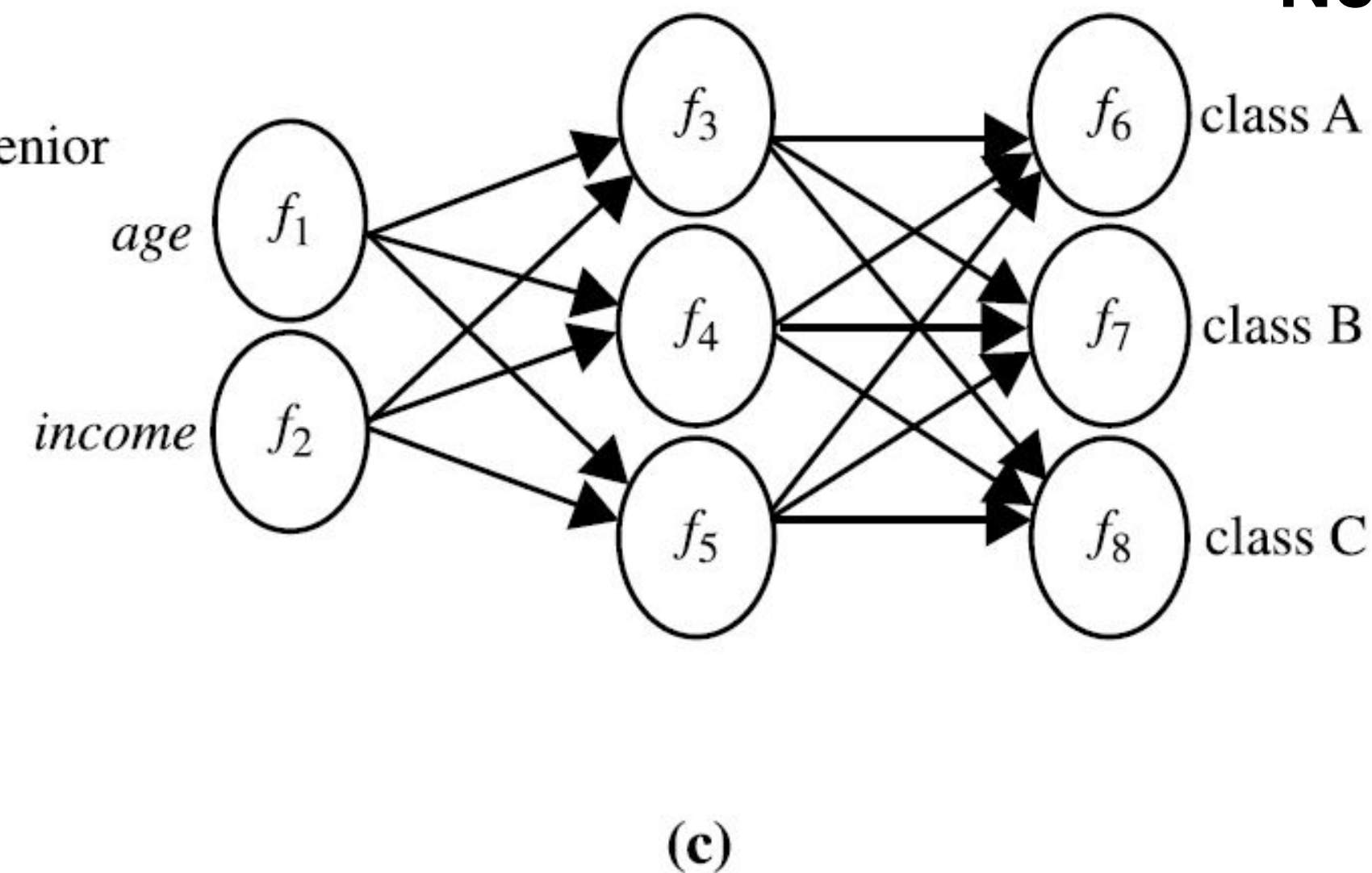
$age(X, \text{"youth"}) \text{ AND } income(X, \text{"high"}) \longrightarrow class(X, \text{"A"})$   
 $age(X, \text{"youth"}) \text{ AND } income(X, \text{"low"}) \longrightarrow class(X, \text{"B"})$   
 $age(X, \text{"middle\_aged"}) \longrightarrow class(X, \text{"C"})$   
 $age(X, \text{"senior"}) \longrightarrow class(X, \text{"C"})$

(a)

## Decision trees



## Neural networks





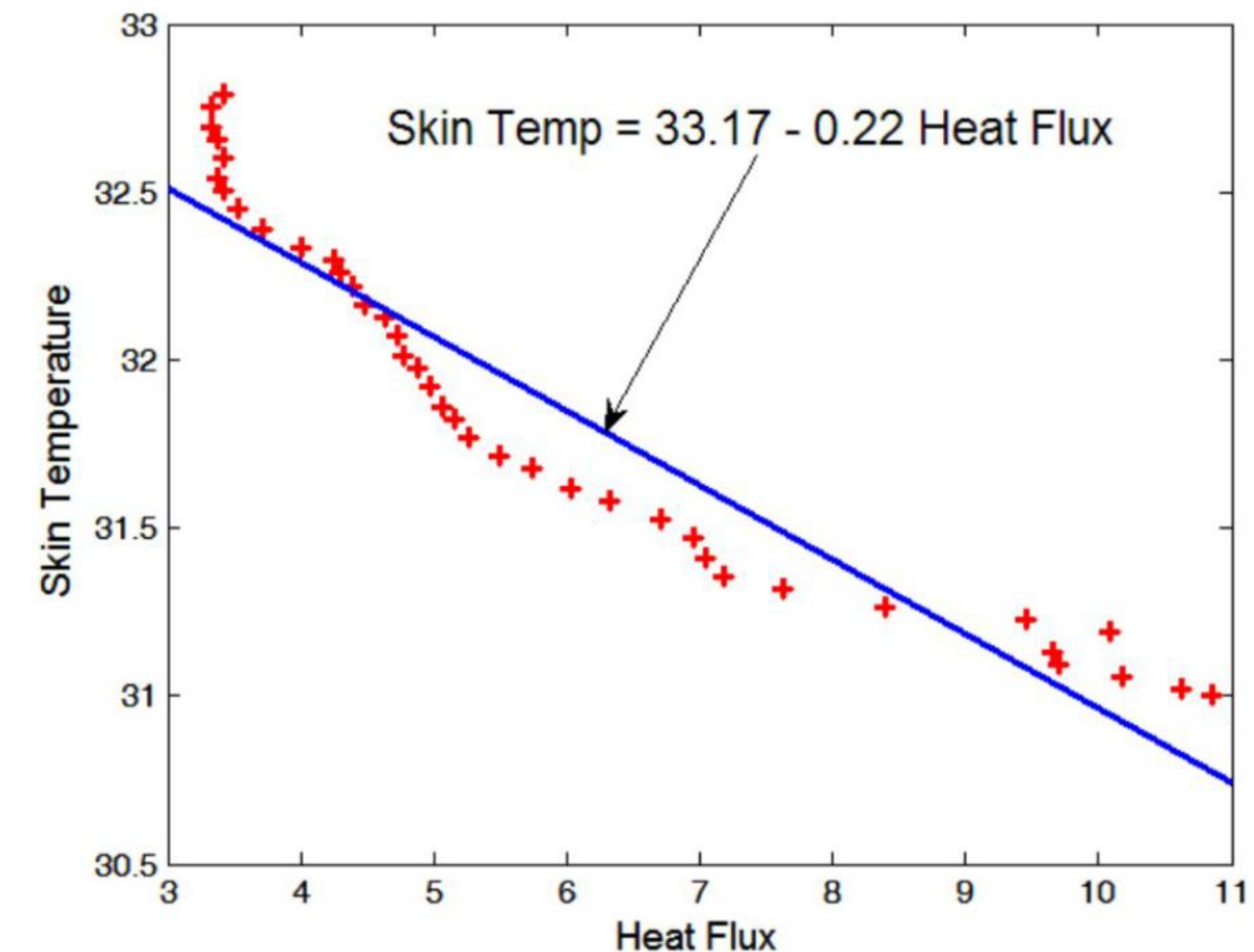
# Predictive Modeling: Regression

Classification predicts **symbolic** (categorical, discrete, unordered, nominal) labels.

Regression predict missing or unavailable **numerical** data values.

Heat Flux	Skin Temperature	Heat Flux	Skin Temperature
10.858	31.002	6.3221	31.581
10.617	31.021	6.0325	31.618
10.183	31.058	5.7429	31.674
9.7003	31.095	5.5016	31.712
9.652	31.133	5.2603	31.768
10.086	31.188	5.1638	31.825
9.459	31.226	5.0673	31.862
8.3972	31.263	4.9708	31.919
7.6251	31.319	4.8743	31.975
7.1907	31.356	4.7777	32.013
7.046	31.412	4.7295	32.07
6.9494	31.468	4.633	32.126
6.7081	31.524	4.4882	32.164

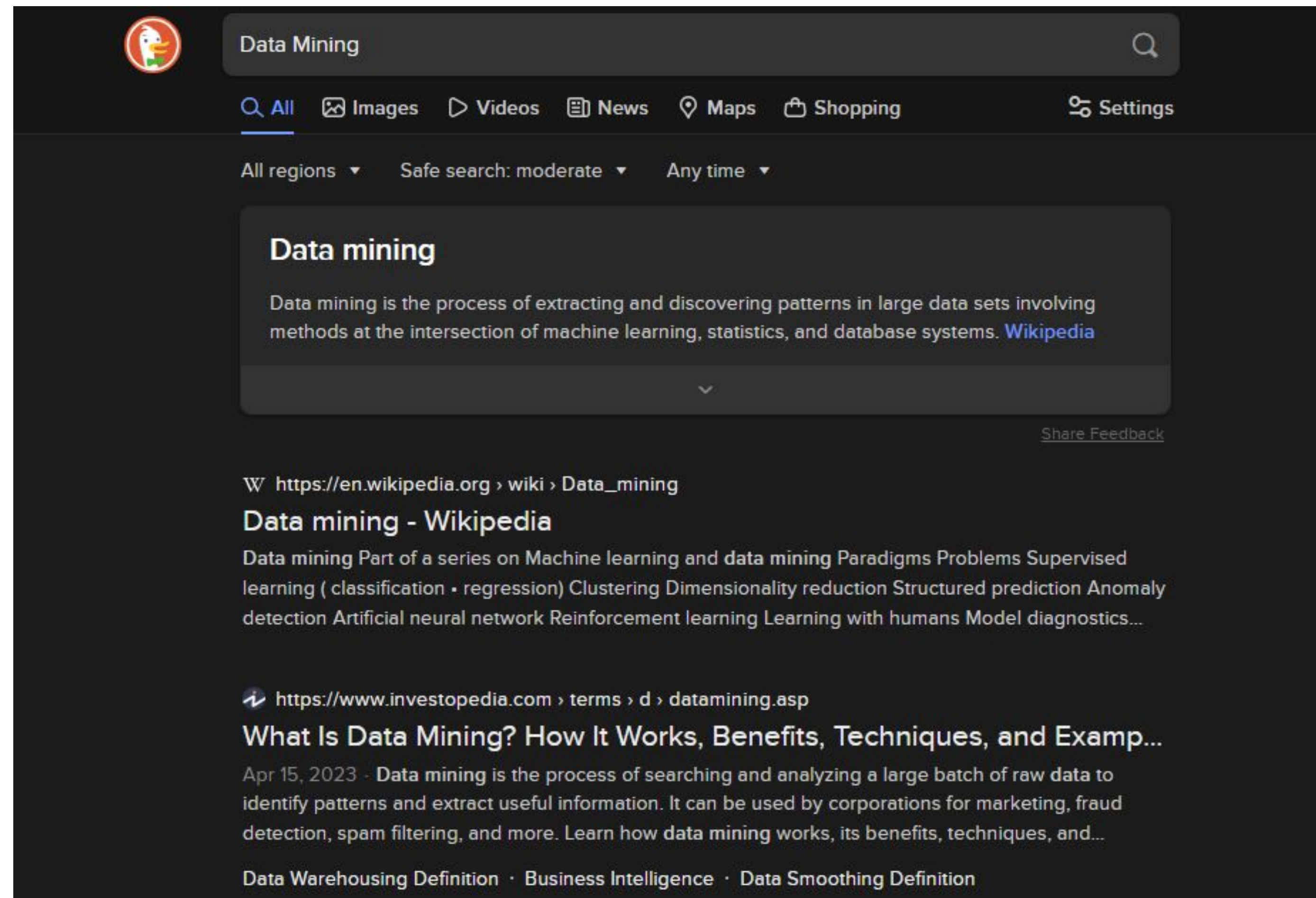
Example: Physiological data from wearable device





# Predictive Modeling: Ranking

The target attribute to be predicted is **ordinal**-valued.



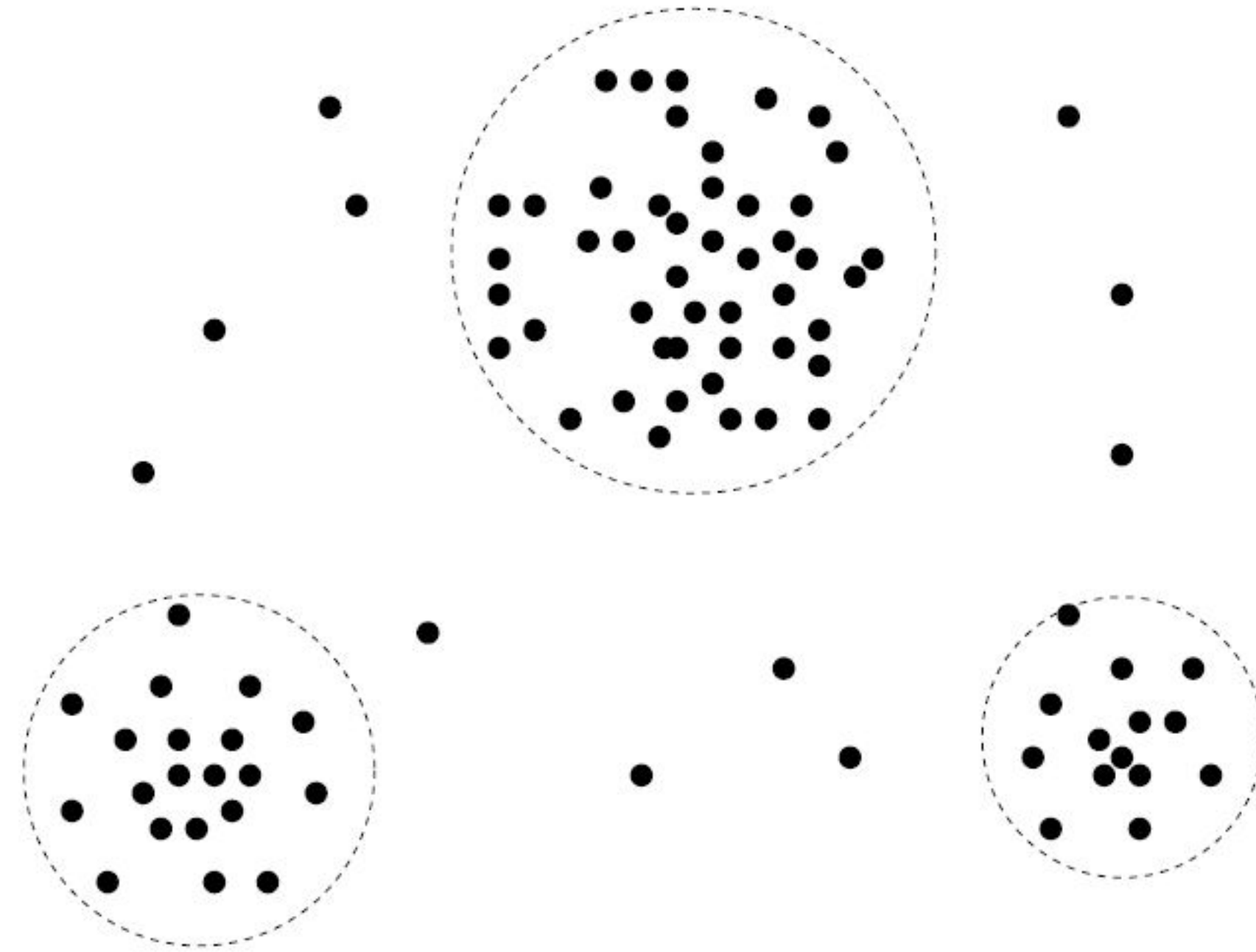


# Clustering

**Goal:** Find groups of objects such that the objects in the same group are more similar to each other than objects from other groups.

Examples:

- Document clustering
- Time series clustering
- Phylogeny reconstruction

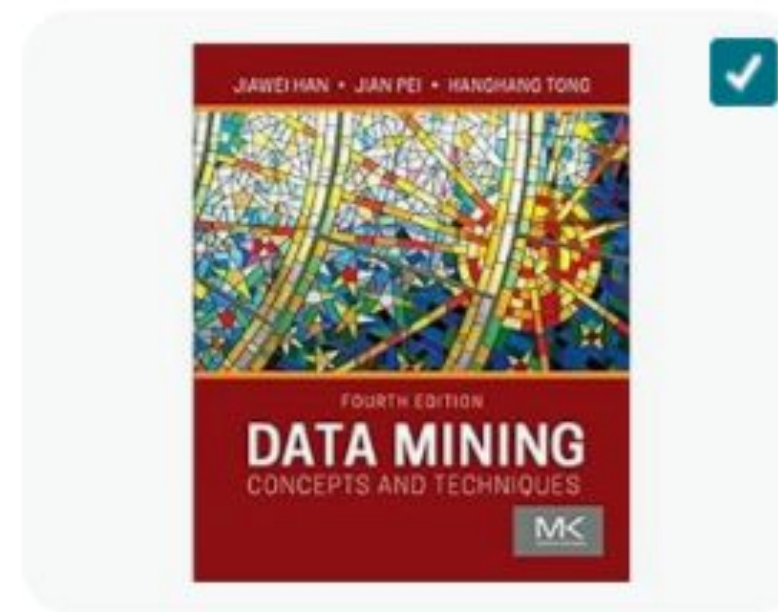


# Association Rule Mining

**Goal:** Find associations (e.g., relationships, dependencies) in sets of data items.

**What items are frequently purchased together in your Amazon transactions?**

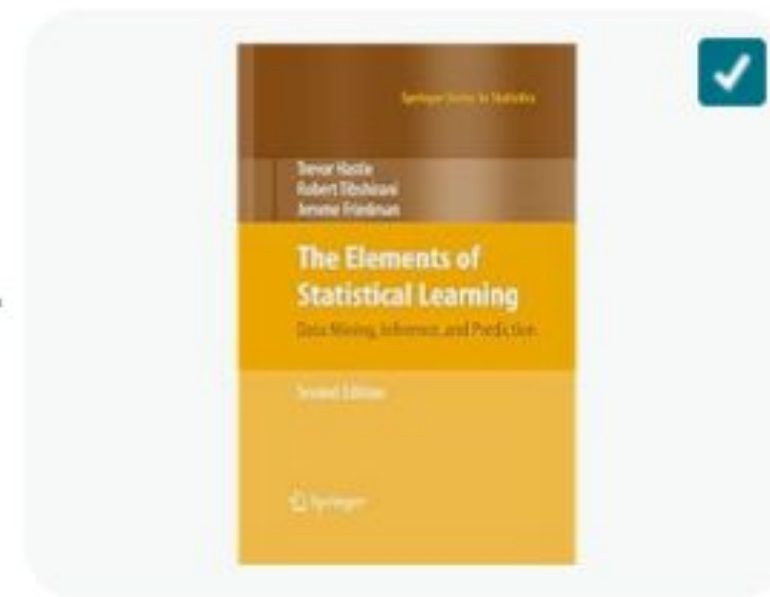
Frequently bought together



**This item:** Data Mining:  
Concepts and Techniques  
(The Morgan Kaufmann...)

\$57<sup>99</sup>

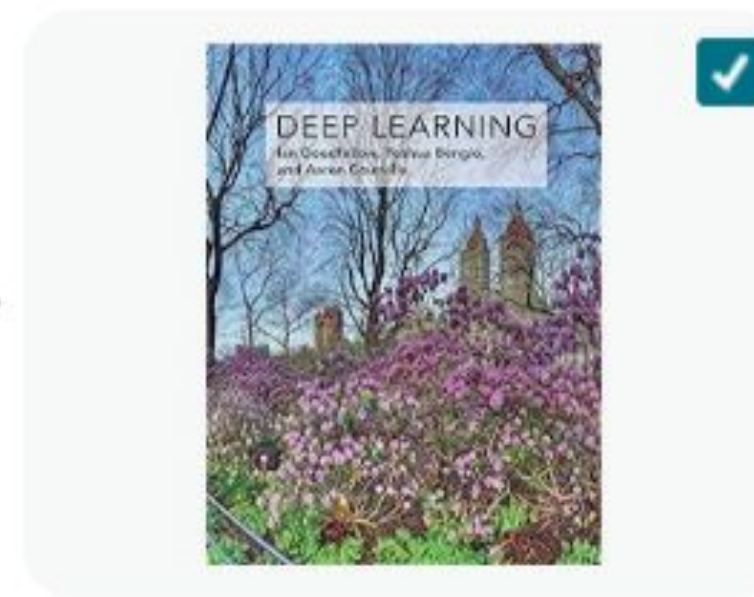
+



The Elements of Statistical  
Learning: Data Mining,  
Inference, and Prediction,...

\$66<sup>13</sup>

+



Deep Learning (Adaptive  
Computation and Machine  
Learning series)

\$93<sup>00</sup>

Total price: \$217.12

Add all 3 to Cart



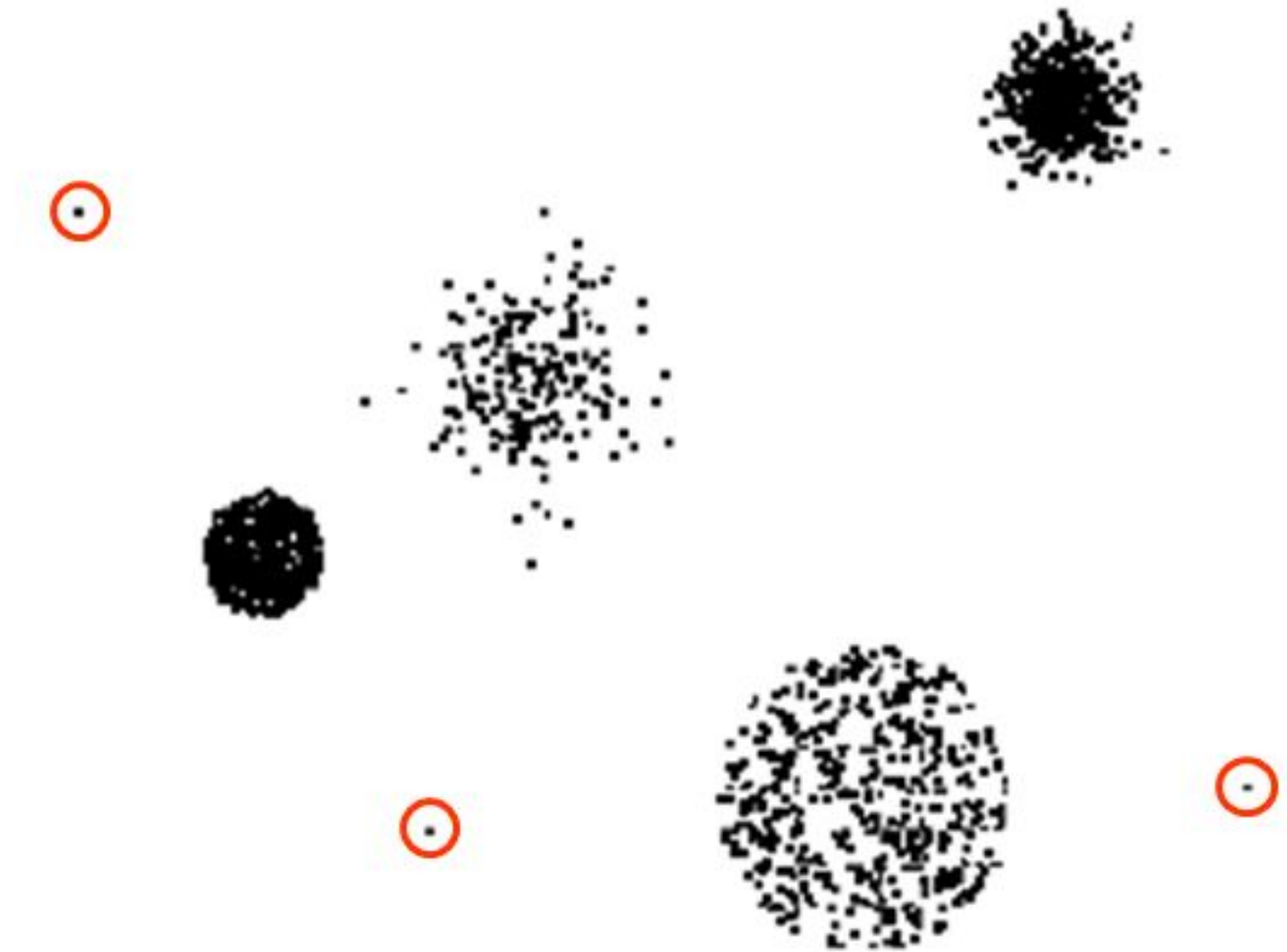
Some of these items ship sooner than the others.  
[Show details](#)

# Anomaly(Outlier) Detection

**Goal:** Identify objects that are different from most other objects.

Examples:

- Credit card fraud detection
- Network intrusion detection
- Event detection in sensor networks
- Defect detection





# Summary

- Common data mining tasks
  - Predictive modeling
    - Classification
    - Regression
    - Ranking
  - Clustering
  - Association rule mining
  - Anomaly detection