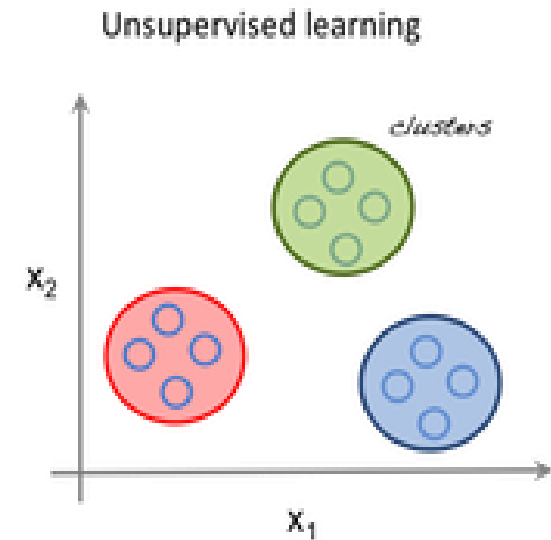
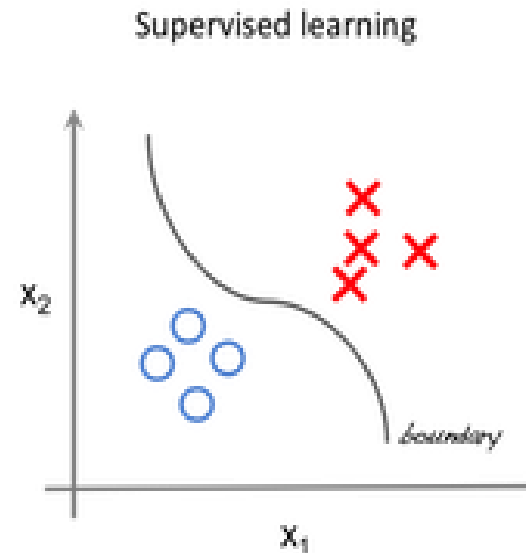


SUPERVISED LEARNING

Supervised learning vs. unsupervised learning

Supervised learning: berusaha untuk menemukan pola pada data yang terkait dengan atribut data tersebut dimana terdapat suatu atribut yang berperan sebagai target.

- Pola yang ditemukan kemudian bisa digunakan untuk tujuan prediksi nilai atribut target pada data baru.



Unsupervised learning: → tidak memiliki atribut target.

- Tujuannya adalah untuk eksplorasi data untuk

Supervised Learning: Uses

Example: decision trees tools that create rules

Prediction of future cases: Use the rule to predict the output for future inputs

Knowledge extraction: The rule is easy to understand

Compression: The rule is simpler than the data it explains

Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

An example application

An emergency room in a hospital measures 17 variables (e.g., blood pressure, age, etc) of newly admitted patients.

A decision is needed: whether to put a new patient in an intensive-care unit.

Due to the high cost of ICU, those patients who may survive less than a month are given higher priority.

Problem: to predict **high-risk patients** and discriminate them from **low-risk patients**.

Another application

A credit card company receives thousands of applications for new cards. Each application contains information about an applicant,

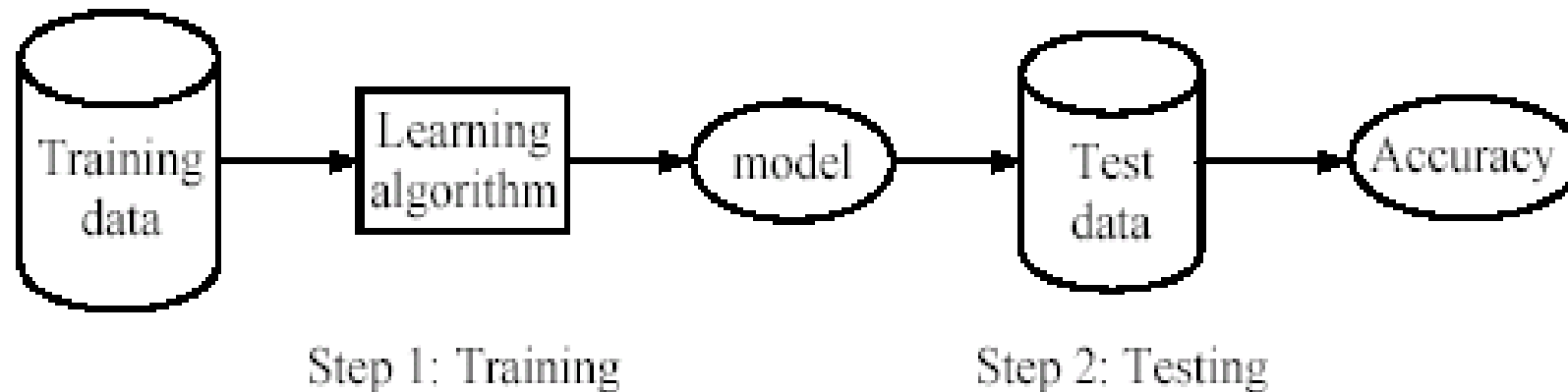
- age
- Marital status
- annual salary
- outstanding debts
- credit rating
- etc.

Problem: to decide whether an application should be approved, or to classify applications into two categories, **approved** and **not approved**.

Supervised learning process: two steps

- **Learning (training):** Learn a model using the training data
- **Testing:** Test the model using **unseen test data** to assess the model accuracy

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$



Supervised Learning techniques

- **Linear classifier** (numerical functions)
- **Parametric** (Probabilistic functions)
 - Naïve Bayes, Gaussian discriminant analysis (GDA), Hidden Markov models (HMM), Probabilistic graphical models
- **Non-parametric** (Instance-based functions)
 - K -nearest neighbors, Kernel regression, Kernel density estimation, Local regression
- **Non-metric** (Symbolic functions)
 - Classification and regression tree (CART), decision tree
- **Aggregation**
 - Bagging (bootstrap + aggregation), Adaboost, Random forest