The data mining process

Elena Baralis
Politecnico di Torino
Data mining

- Non trivial extraction of
  - implicit
  - previously unknown
  - potentially useful
  information from available data
- Extraction is automatic
  - performed by appropriate algorithms
- Extracted information is represented by means of abstract models
  - denoted as *pattern*
Knowledge Discovery Process

KDD = Knowledge Discovery from Data
Analysis techniques

- Descriptive methods
  - Extract interpretable models describing data
  - Example: client segmentation

- Predictive methods
  - Exploit some known variables to predict unknown or future values of (other) variables
  - Example: “spam” email detection
Analysis techniques

- Association rule extraction

- Classification

- Clustering