

Introduction --- Concepts

Data mining is the process of discovering patterns in data.

For what purpose? To make nontrivial predictions on new data.

Table 1.1--- all possible data is presented and it is possible to summarize the data as a simple rule---this is not always the case

Real life data---incomplete, unknown fields, noisy data

Attributes and outcome

Symbolic (nominal) vs. numeric (ordinal) values

Classification rules --- predict the classification of the outcome

Association rules --- associate different attribute values

Decision list (list of rules where ordering is important)

Decision tree

Numeric prediction --- regression equation

Training data and test data

Overfitting the training data

Supervised learning --- each training data has an associated classification

Unsupervised learning---no such classification is provided; ex: clustering

Training data, test data, new instances of data

ARFF Attribute-Relation File Format) format

Standardizing numerical attributes (in the input)

Output: Decision tables, decision trees, (Divide-and-conquer approach), classification rules (antecedent and consequent), association rules, rules with exceptions, rules involving relations (i.e., relation among the attributes)