

Data Mining

Spring 2013

Data mining is the process of discovering meaningful correlations, patterns, regularities in data consisting of many records and variables. Often little is known about the distribution of any particular variable, or about the relationships between variables. Data mining is part of more general process called Knowledge Discovery: extracting useful knowledge from raw data.

Topics include:

- Data Preparation and Cleaning
- Association Rules
- Logistic Regression
- Classification Trees
- Regression Trees
- Simpson's Paradox
- Nearest Neighbor Rule
- Neural Nets
- K-means Clustering
- Hierarchical Clustering
- Exploratory Data Analysis and Visualization
- Receiver Operating Curves
- False Discovery Rates