Datamining

- artificial intelligence, databases, statistics

- one view: a step in the overall process of knowledge discovery from data (KDD)
  - KDD: finding understandable patterns that can be interpreted
  - modeling and supporting human-machine interaction

- steps of KDD (iterative and interactive)
  - learn the application domain
  - create a target dataset (on which discovery performed)
  - pre-process (e.g. handle missing data fields)
  - data reduction and projection (find useful features)
  - choose function for data mining (e.g. classification)
  - choose data mining algorithm
  - do data mining
  - interpretation of discovered patterns (form for users)

- datamining steps
  - the model: its function and representation (show parameters to be determined from data)
  - the preference criterion: goodness of fit
  - the search algorithm
  - often all mixed up in 1 learning algorithm

- various challenges
  - high performance: rapid response to assist users
  - understandability of patterns (e.g. graphical representations)
  - interpretation: standalone not that useful; perhaps integrate with query interface
  - accommodate data types beyond numeric: e.g. multimedia