CSE 574: Artificial Intelligence II

Data Mining and Machine Learning

Instructor:

Pedro Domingos

Logistics

- **Instructor:** Pedro Domingos
  Email: pedrod
  Office: Sieg 216
  Office hours: Thu 11:00-12:00 or by appointment
- **TA:** Geoff Hulten
  Email: ghulten
  Office: Chateau 104
- **Web:** www.cs.washington.edu/574
- **Mailing list:** cse574

Assignments

- **Project**
  - Individual or in groups of two
  - From list or proposed by students
  - Class grade = Project grade
- **Seminar**
  - May be related to project or not
  - Pass/Fail

Source materials

- Tom Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy (editors), *Advances in Knowledge Discovery and Data Mining*, AAAI Press/MIT Press, 1996.
- David Hand, Heikki Mannila, and Padhraic Smyth, *Principles of Data Mining*, MIT Press, to appear.
- Papers

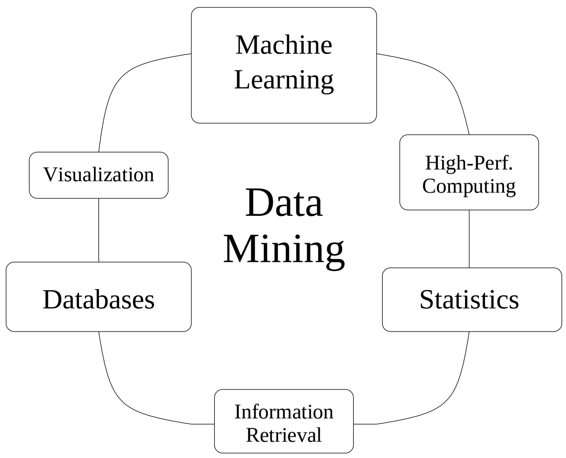
What is data mining?

Data mining is the process of identifying valid, novel, useful and understandable patterns in data.

Also known as **KDD** (<u>K</u>nowledge <u>D</u>iscovery in <u>D</u>atabases).

"We're drowning in information, but starving for knowledge."

John Naisbett



Data mining as an AI-complete problem

- Machine learning
- Natural language
- Vision
- Speech
- Planning
- Agents
- Knowledge representation & reasoning
- Common sense

Applications of data mining

Current

- Marketing and retail
- Finance
- Telecoms
- World Wide Web
- Drug design
- Process control
- Space and earth sensing

Breaking

- E-commerce
- Software
- Health care
- Second-wave reengineering
- Adaptive environments
- Anywhere there's data?

The data mining process

- Understanding domain, prior knowledge, and goals
- Data integration and selection
- Data cleaning and preprocessing
- Modeling and searching for patterns
- Interpreting results
- Consolidating and deploying discovered knowledge
- Loop

Data mining tasks

- Classification
- Regression
- Probability estimation
- Clustering
- Association detection
- Summarization
- Trend and deviation detection

Inductive learning

Inductive learning:

Given examples of a function $\langle X, F(X) \rangle$

Predict function $F(X)$ for new examples X

- Discrete $F(X)$: Classification
- Continuous $F(X)$: Regression
- $F(X) = \text{Probability}(X)$: Probability estimation

Inductive learning (contd.)

Widely-used approaches:

- Decision trees
- Rule induction
- Instance-based learning (kNN)
- Neural networks
- Bayesian networks
- Etc.

Requirements for a data mining system

Data mining systems should be

- Computationally sound
- Statistically sound
- Ergonomically sound

Components of a data mining system

- Representation
- Evaluation
- Search
- Data management
- User interface

Topics for this quarter

- Classification (DTs, rules, kNN)
- Regression
- Model ensembles
- Probability estimation
- Clustering
- Association rule mining
- Scaling up
- Data preparation
- Postprocessing

Projects: Factors to consider

- How well-defined?
- How open-ended?
- Potential impact?
- How much math & statistics?
- How much implementation work?
- Data available?
- Expertise available?

Project ideas

- Extending the scope of learning algorithms
- Web mining
- Scaling up
- Using frequent sets for prediction
- Overfitting avoidance
- Model ensembles
- Learning with prior knowledge

Extending the scope of learning algorithms

- Learning time-changing concepts
- Learning given a query distribution
- Learning with thousands of classes
- Learning from non-random samples
- Learning from multiple databases
- Learning from mixed-media data
- Learning across populations

Scaling up

- Adaptive sampling
- Scalable data squashing
- Interleaving learning and compression

Overfitting avoidance

- Estimating the level of noise in a database
- Learning noise models
- Empirical process-oriented evaluation of decision trees
- Knowledge-intensive evaluation functions

Web mining

- Modeling the hidden Web through querying
- Learning to place Web pages in Yahoo! categories
- Clickstream mining: From sequences to shortcuts
- Web dynamics: Predicting changes in access frequency

Using frequent sets for prediction

- Extending naive Bayes with frequent sets
- Frequent sets as features for large-margin classifiers
- Probability distributions from frequent sets via iterative scaling
- Learning Bayes nets via frequent sets

Model ensembles

- Combining multiple decision trees into one
- Avoiding overfitting in boosting
- Scaling up alternating decision trees

Learning with prior knowledge

- Learning with weak causal knowledge
- Learning to correct input errors
- Online handwriting recognition