CSCI 1950-F: Introduction to Machine Learning

Brown University, Summer 2011

How can artificial systems learn from examples, and discover information buried in massive datasets? This course explores the theory and practice of statistical machine learning. Topics include parameter estimation, Bayesian inference, probabilistic graphical models, approximate inference, and kernel methods. Applications to regression, classification, and clustering problems are illustrated by examples from vision, language, communications, and bioinformatics. *Prerequisites:* Programming experience, and comfort with basic probability, linear algebra, and calculus.

Introduction

The main goal of this class is to introduce you to the ideas and techniques of machine learning, and the probabilistic models that underlie behind them. These ideas have their origins in classical results from statisticians such as Laplace, Bayes, and Fisher. However, modern computing techniques now permit applications of a scale and diversity that was barely conceivable only a few decades ago.

As opposed to the traditional statistical focus on analysis of experiments, most problems we'll discuss involve some form of prediction. *Classification* algorithms predict a discrete value from a finite set of choices, while *regression* algorithms predict a continuous value. *Supervised learning* techniques can be used to design such predictors using training data that is labeled with the values you are trying to learn. *Unsupervised learning* techniques are instead used when such labels are unavailable, but you nevertheless hope to discover interesting structure within your data. These methods lead to effective algorithms for *clustering* and *dimensionality reduction*. This course will explore the conceptual relationships between these different learning problems, and introduce some of the most practically effective statistical models and computational methods.

Administrative Information

Lecture Time/Location: The class will meet each weekday (Monday through Friday) from 10:30 am - 12:00 noon, in CIT room 345 (at 115 Waterman Street).

Instructor:
Jeff Miller (jeffrey_miller@brown.edu, 484-459-0587)

Grading

Overall grades will be assigned as follows: 40% homeworks, 40% quizzes, 20% class participation. Each day, there will be a short homework assignment and a short quiz.

Course website

See the course website at
`http://www.dam.brown.edu/people/jmiller/ML/index.html`
for more information.

Schedule outline

See the course website (above) for a detailed schedule.

Week 1:

- Introduction, Probability review, K-nearest neighbor classification
- Trees
- Decision theory
- MLE and MAP estimation
- Bayesian inference

Week 2:

- Naïve Bayes classification
- Linear regression
- Multivariate Gaussian distribution
- Bayesian linear regression
- Estimators
- Model selection

Week 3:

- Directed graphical models ("Bayesian" networks)
- Hidden Markov models (HMMs), Dynamic programming
- Logistic regression, and Newton's method (for optimization)
- Clustering with K-means and Gaussian mixture models (GMMs)
- Expectation-Maximization (EM) algorithm

Week 4:

- Sampling, Monte Carlo approximation, Importance sampling
- Markov chain Monte Carlo (MCMC)
- Sparse kernel methods
- Semi-supervised learning (SSL)
- Dimensionality reduction