CSCI 1950-F: Introduction to Machine Learning

Erik Sudderth and Mark Johnson, Fall 2009

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, probabilistic graphical models, approximate inference, and kernel and nonparametric methods. Applications to regression, categorization, and clustering problems are illustrated by examples from vision, language, communications, and bioinformatics. *Prerequisites:* CSCI0160, CSCI0180 or CSCI0190, 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 work by statisticans such as Laplace and Bayes several centuries ago. However, modern computing techniques now permit us to apply these to problems of a size and diversity that was barely conceivable only a few decades ago.

The kinds of problems we'll discuss involve prediction of one kind or another. *Classification* problems involve predicting a discrete value from a finite set of choices, while *regression* problems involve predicting a continuous value. *Supervised learning* techniques can be used to design such predictors from 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, as well as introduce some of the most practically effective computational methods.

Administrative Information

Lectures: Tuesdays and Thursdays from 2:30-4:00pm, CIT room 227, 115 Waterman St.

Instructors:

Erik Sudderth (`sudderth@cs.brown.edu`; 401-863-7660)
Office Hours: Mondays 4pm-6pm, CIT room 509, 115 Waterman St.

Mark Johnson (`Mark_Johnson@brown.edu`; 401-863-1670)
Office Hours: Wednesdays 10am-noon, Metcalf Research Bldg. room 231, 190 Thayer St.

Graduate Teaching Assistant:

Deqing Sun (`dqsun@cs.brown.edu`)
Office Hours: Tuesdays 8pm-10pm, CIT room 227, 115 Waterman St.

Undergraduate Teaching Assistants:

Maximilian Barrows (`mbarrows@cs.brown.edu`)
Office Hours: Wednesdays 10pm-midnight, CIT room 227, 115 Waterman St.

Evan Donahue (`emdonahu@cs.brown.edu`)
Office Hours: Wednesdays 3pm-5pm, CIT room 227, 115 Waterman St.

Grading, Assignments, and Readings

There will be between eight and ten homework assignments, each due at least one week after it is assigned. Homework problems will involve both mathematical derivations and Matlab implementation of learning algorithms. Electronic submission instructions, and a formal collaboration policy, will be announced with the first assignment on September 17. **Late submissions will not be accepted without prior approval.**

In addition to homeworks, there will be one in-class midterm on Oct. 22, and a final exam. Overall grades will be assigned as follows: 50% homeworks, 25% final exam, 20% midterm exam, 5% class participation. Graduate students may receive 200-level credit by solving additional, more advanced homework and exam problems. These graduate-level problems will typically involve more sophisticated mathematics, as well as more challenging generalizations of the lecture material.

There is no required textbook, but we strongly recommend Christopher Bishop's 2006 *Pattern Recognition and Machine Learning* text. We will provide supplemental readings from this book, and occasionally other sources, for each lecture.

Tentative Syllabus

The details of this schedule are almost certain to change!

09/10: Introduction to machine learning problems and applications
09/15: Binary and multinomial classification, discrete naive Bayes, ROC curves
09/17: Probability review: Bayes rule, directed graphical models, maximum likelihood estimation
09/22: K-nearest neighbors classification, continuous naive Bayes, cross-validation
09/24: Cross-validation, model selection (AIC/BIC), linear least squares regression
09/29: Linear algebra and the geometry of multivariate Gaussian distributions
10/01: Least squares regression, Bayesian and frequentist estimation, MAP
10/06: L2 versus L1 regularization, lasso, logistic regression
10/08: ML/MAP optimization: gradient descent, stochastic gradient, quasi-Newton methods
10/13: Perceptron algorithm, kernels, and margins
10/15: Support vector machines, kernels, and margins
10/20: Gaussian process regression and classification
10/22: *Midterm Exam*
10/27: Graphical models: directed versus undirected, Markov properties, examples
10/29: Clustering: K-means/medoids, convergence, initialization
11/03: Expectation-Maximization (EM) algorithm for clustering
11/05: Variational interpretation of the EM algorithm
11/10: Introduction to hidden Markov models (HMMs), Viterbi algorithm
11/12: Sum-product algorithm, EM for HMM parameter estimation
11/17: Dimensionality reduction: PCA, factor analysis
11/19: Monte Carlo estimators, importance sampling, sequential Monte Carlo
11/24: MCMC: Gibbs samplers, Metropolis-Hastings
11/26: *Thanksgiving*
12/01: Vision and robotics: state space models, switching dynamical systems, neural networks
12/03: Natural language processing: topics models, probabilistic grammars
12/07: Reading week, review for upcoming final exam.