

CSCI 1420: Machine Learning

ENGN 2520: Pattern Recognition and Machine Learning

Brown University, Fall 2015

How can artificial systems learn from examples, and discover information buried in massive datasets? We explore the theory and practice of statistical machine learning, focusing on computational methods for supervised and unsupervised data analysis. Specific topics include Bayesian and maximum likelihood parameter estimation, regularization and sparsity-promoting priors, kernel methods, the expectation maximization algorithm, and models for data with temporal or hierarchical structure. Applications to regression, categorization, clustering, and dimensionality reduction problems are illustrated by examples from vision, language, bioinformatics, and information retrieval.

Prerequisites: Comfort with basic multivariable calculus, an introductory programming course (CSCI0040, CSCI0150, CSCI0180, or CSCI0190), an introductory probability course (CSCI1450, APMA1650, or MATH1610), and an introductory linear algebra course (CSCI0530, MATH0520, or MATH0540); or permission of instructor.

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

Textbooks: *Pattern Recognition and Machine Learning*, C. M. Bishop, Springer 2007.

Elements of Statistical Learning, T. Hastie, R. Tibshirani, & J. Friedman, Springer 2009.

Lectures: Tuesdays and Thursdays from 2:30-3:50pm, Smith-Buonanno Hall 106.

Recitations: Monday afternoons, time and location TBD. Led by the graduate teaching assistant.

Instructor: Prof. Erik Sudderth (sudderth@cs.brown.edu; 401-863-7660; CIT room 555)

Graduate Teaching Assistant: Geng Ji (gji@cs.brown.edu; CIT room 545)

Undergraduate Teaching Assistants: Emily Smith (Head UTA), Angelia Wang, Christopher Grimm, Karthik Harihar Reddy Battula, Keshav Vemuri, Sarah Sachs, & Zachary Nado

Office Hours: See the course website for a detailed schedule.

Homework Assignments

There will be nine homework assignments, each due at least one week after it is handed out. Homework problems will combine mathematical derivations and Matlab implementation of learning algorithms. The scores of all nine assignments will be averaged equally to determine an overall score (we will not “drop” any homeworks). Homeworks will be submitted electronically.

Collaboration Policy Students may discuss and work on homework problems in groups. However, each student must write up their solutions independently, and do any required programming independently. List the names of any collaborators on the front page of your solutions. You may *not* directly copy solutions from other students, or from materials distributed in previous versions of this or other courses. You may *not* make your solutions available to other students: files in your home directory may not be world-readable, and you may not post your solutions to public websites.

Late Submission Policy Homework assignments are due by 11:59pm on Thursday evenings. Your answers may be submitted up to 4 days late (by Monday evening); after this point, solutions will be distributed and handins will no longer be accepted. You may submit up to two late assignments without penalty. For each subsequent late assignment, 20 points (out of a maximum of 100) will be deducted from the overall score. *Exceptions to this policy are only given in very unusual circumstances, and any extensions must be requested in advance by e-mail to the instructor.*

Exams and Course Grades

In addition to homeworks, there will be one take-home midterm exam, and a take-home final exam. You will have three days to complete each exam. You may not discuss exam questions with other students, and late submissions of midterm and final exams will not be accepted. Overall course grades will be assigned as follows: 65% homeworks, 15% midterm exam, 20% final exam. Undergraduate and graduate teaching assistants contribute to the grading of course assignments.

Syllabus: Summary of Course Topics

Classification problems ROC curves, decision theory, loss functions

Classification models naïve Bayes, nearest neighbors, decision trees, neural networks

Frequentist learning maximum likelihood (ML) estimation, asymptotics

Bayesian learning Bayes rule, MAP and MMSE estimation, marginal likelihood

Linear models multivariate Gaussian distributions, linear regression, logistic regression

Model selection validation and cross-validation, regularization and priors

Feature selection penalized likelihoods, search, sparsity and L_1 regularization

Learning via optimization (stochastic) gradient descent, quasi-Newton methods, boosting

Kernel methods Gaussian processes, support vector machines (SVMs)

Clustering K-means, mixture models, expectation maximization (EM) algorithm

Dimensionality reduction PCA, factor analysis, manifold learning

Hidden Markov models (HMMs) Viterbi and sum-product algorithms, EM algorithm

Directed graphical models Markov properties, exponential families, applications