

IEMS 490

Introduction to Machine Learning

Mo-Wed: 11-12:20, Tech MG28

Professor: Jorge Nocedal, Office M326 Tech

Teaching Assistant: Stefan Solntsev, Office L375 Tech
stefans@u.northwestern.edu

Course Outline

Machine learning techniques are used with great success in important application areas, such as computer vision, search engines, speech recognition, robotics, recommendation systems, bioinformatics, social networks, and finance. This course provides an introduction to machine learning, with emphasis on optimization methods as learning algorithms. The course begins by introducing several examples of supervised and unsupervised learning. The main body of the course focuses on the design of statistical learning models and on the optimization algorithms that are used to train them.

Syllabus: (See the end of the document for lecture by lecture breakdown of the course material.)

Overview

What is machine learning? The three pillars: feature selection, statistical modeling, learning via optimization. Examples: text classification; speech recognition; recommendation systems.

Supervised Learning

Linear regression. The stochastic gradient descent method of Robbins-Monro. Training error and generalization error

Higher-order models, over-fitting and L1 regularization

Locally weighted linear regression.

Classification and logistic regression. Stochastic gradient descent method for logistic regression.

Generalized linear models. The exponential family. Softmax regression.

Generative learning algorithms. Gaussian discriminant analysis. Naïve Bayes approach.

Stochastic averaging optimization (variance reduction)

Support vector machines (definition and application)

Introduction to convex optimization: duality

Support vector machines (derivation and analysis)

Optimization methods for training support vector machines

Model selection. Feature selection

Boosting

Learning theory

The perceptron algorithm

VC dimension. Bias vs variance

Unsupervised Learning

Clustering. K-means.

The EM Method. Mixture of Gaussian

PCA and IPCA (brief summary; assume students know topic)

Deep Learning (Modern Neural Networks)

Motivation, definition and training

Prerequisites:

Good working knowledge of a programming language (Matlab is fine); strong foundations in statistics and probability (at the undergraduate level) and linear algebra. Knowledge of nonlinear optimization methods and theory is *not* required (but is certainly useful).

Grading

Homeworks 60%, project 40%. Students can select the emphasis of their projects: an application area, a study of machine learning models, or an optimization method.

References:

The course follows the general organization of the class taught by Andrew Ng at Stanford <http://cs229.stanford.edu/>

The following textbook is recommended as a reference: Christopher Bishop, “Pattern Recognition and Machine Learning” Springer 2006. Some lectures will be based on this book.



Lecture by Lecture:

March 31 What is machine learning? Features, labels, cost functions. The stochastic gradient descent method of Robbins-Monro. Optimization as a learning process. Training error vs Generalization error, over-fitting and L1 regularization. An example of a cutting edge research problem: speech recognition (SIRI vs Google vs Microsoft)

April 1 Logistic Regression. The stochastic gradient method for logistic regression and how it suggests the concept of generalized linear models. The perceptron algorithm. A few more comments about L1 regularization.

April 7 Multi-class logistic classification. Online Learning. Introduction to Neural Networks. Locally weighted linear regression

April 9 Support vector machines. Separable case. Properties of the gradient method

April 14 Duality, Support Vector

April 16 Kernels. Properties of the stochastic gradient method.

April 21 Generative Learning methods. Naïve Bayes method