CSE6740/CS7641/ISYE6740: Machine Learning I
Fall 2012

Lecture Time

Tuesday and Thursday 1:35 - 2:55pm in Klaus 2447 (starting Aug 21st)

Course Description

Machine learning studies the question "how can we build computer programs that automatically improve their performance through experience?" This includes learning to perform many types of tasks based on many types of experience. For example, it includes robots learning to better navigate based on experience gained by roaming their environments, medical decision aids that learn to predict which therapies work best for which diseases based on data mining of historical health records, and speech recognition systems that lean to better understand your speech based on experience listening to you.

The course is designed to answer the most fundamental questions about machine learning: How can we conceptually organize the large collection of available methods? What are the most important methods to know about, and why? How can we answer the question 'is this method better than that one' using asymptotic theory? How can we answer the question 'is this method better than that one' for a specific dataset of interest? What can we say about the errors our method will make on future data? What's the 'right' objective function? What does it mean to be statistically rigorous? Should I be a Bayesian? What computer science ideas can make ML methods tractable on modern large or complex datasets? What are the open questions?

This course is designed to give PhD students a thorough grounding in the methods, theory, mathematics and algorithms needed to do research and applications in machine learning. The course covers topics from machine learning, classical statistics, data mining, Bayesian statistics and information theory. Students entering the class with a pre-existing working knowledge of probability, statistics and algorithms will be at an advantage, but the class has been designed so that anyone with a strong numerate background can catch up and fully participate.

If a student is not prepared for a mathematically rigorous and intensive class of machine learning, I suggest you take: Data and Visual Analytics course in the Spring, CSE 6242.

Textbooks

- Pattern Recognition and Machine Learning, Chris Bishop
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Trevor Hastie, Robert Tibshirani, Jerome Friedman.
- Machine Learning, Tom Mitchell

Grading

The requirements of this course consist of participating in lectures, final exam, 5 problem sets and a project. This is a PhD level class, and the most important thing for us is that by the end of this class students understand the basic methodologies in machine learning, and be able to use them to solve real problems of modest complexity. The grading breakdown is the following

- Homework (5 assignments, 60%)
- Final exam (20%)
- Final project (20%)

Exams

The midterm and final exams will be open book and open notes. Computers will not be allowed.

People

Instructor: Le Song, Klaus 1340, Office Hours: Thursday 3-4pm

Guest Lecturer: TBD

TA Office Hours: Monday 3-4pm and Thursday 3-4pm

TA: Seungyeon Kim, Klaus 1305

TA: Tran Quoc Long, Klaus 1305

TA: Parikshit Ram , Klaus 1305

Class Assistant: Michael Terrell, Klaus 1321

Mailing List

Discussion forum: https://groups.google.com/d/forum/cse6740fall2012

Mailing list: cse6740fall2012@googlegroups.com

Email to contact TA: cse6740.fall2012@gmail.com

Syllabus and Schedule

Date	Lecture & Topics	Readings & Useful Links	Handouts
Tue 8/21	Lecture 1: Introduction • What is Machine Learning? • Applications of Machine Learning • Basic Machine Learning Models • Logistics	• Science special issue on data science • Nature special issue on big data • Reading: Biship book: Chapter 1, 2 • Andrew Moore's tutorials	Slides
Introduction to Functional Approximation: Density Estimation (eg. Maximum likelihood principle, Overfitting, Bayesian versus Frequentist estimate), Classification Theory, Optimal Classifier, Nonparametric methods & Instance-based learning (eg. Bayesian decision rule, Bayes error, Parzen and nearest neignbor density estimation, K-nearest neighbor (kNN) classifier)			

Thu 8/23	Lecture 2: Classification Theory, Optimal Classifier
	• Nonparametric methods & Instance-based learning
	  ◦ Bayesian decision rule
	  ◦ Parzen and nearest neignbor density estimation
		• Biship book: Chapter 2.5
		• Fukunaga book (Introduction to Statistical Pattern Recognition):
		  • hypothesis test
		  • nonparametric density estimation
		  • nonparametric
			Slides

	◦ K-nearest neighbor (kNN) classifier	classification	
Linear Function Learning: Naive Bayes classifiers, Linear Regression, Logistic regression, Discriminative v.Generative models			
Tue 8/28	Lecture 3: Generative classifiers • Naive Bayes classifiers with discrete and continuous (Gaussian) features ◦ Applet	• Reading: • Naive Bayes and Logistic Regression, Mitchell book chapter draft • Bishop book, Chapter 4	Slides
Thu 8/30	Lecture 4: Discriminative classifiers: • Linear rregression • Its probabilistic interpretation • Applet	• Reading: • Mitchell book chapter 8.3 • Bishop book, Chapter 3	Slides
Tue 9/4	Lecture 5: Discriminative classifiers: • Logistic regression • Relationship to Naive Bayes • Applet	• Reading: • Mitchell book chapter 4 • Bishop book, Chapter 4, 5 • Naive Bayes and Logistic Regression, Mitchell book chapter draft • On Discriminative and Generative Classifiers, Ng and Jordan, NIPS, 2001	Slides

Non-Linear Models and model selection: Decision trees, Neural networks, Support Vector Machines, Kernel Methods, Boosting			
Tue 9/6	Lecture 6: Complex discriminative function learning • Decision tree learning • Applet	• Reading: • Mitchell book chapter 3 on decision trees • Bishop book, Section 1.6	Slides
Tue 9/11	Lecture 7: Neural Networks • Non-linear regression, classifiers • Gradient descent • Discovered representations at hidden layers • Applet	• Reading: • Bishop, Chap 5 • Mitchell, Chap 4	Slides
Tue 9/13	Lecture 8: Support Vector Machine • Duality • Kernel Methods • Convex Optimization	• Reading: • Bishop, Chap 6,7 • Burges tutorial	Slides
Tue 9/18	Lecture 9: Boosting • Combination of clsasifiers • Adaboost • Applet	• Reading: • Bishop, Chap 14.3 • boosting homepage	Slides
Theory and Practice in Supervised Learning: Sample complexity, PAC learning, Error bounds, VC-dimension, Margin-based bounds, Overfitting, Cross validation, Model selection			
	Lecture 10: Learning		

	Tue 9/20	Theory I • Sample complexity • Hypothesis space • Version space • PAC learning theory[Applet] • Agnostic learning	• Reading: • Mitchell book, Chap 7	Slides
	Tue 9/25	Lecture 11: Learning Theory II • VC dimension and Agnostic learning • Overfitting and PAC bounds • Structural Risk Minimization	• Reading: • Mitchell book, Chap 7	Slides
	Tue 9/27	Lecture 12: Practical issues in supervised learning • Overfitting • Bias and variance decomposition • Cross-validation • Regularization	• Reading: • Mitchell book, Chap 5&6 • Bishop, Chap 1&2	Slides
	Tue 10/2	Lecture 13: Introduction to Unsupervised Learning • Clustering • K-means clustering [Applet] • Hierarchical clustering [Applet]	• Reading: • Bishop, Chap 9	Slides
		Lecture 14: Mixture model		

Tue 10/2	• Expectation-Maximization algorithm • [Applet]	• Reading: • Bishop, Chap 9	Slides
Theory and Practice in Supervised Learning: Sample complexity, PAC learning, Error bounds, VC-dimension, Margin-based bounds, Overfitting, Cross validation, Model selection			
Unsupervised Learning and Structured Models: K-means, Expectation Maximization (EM) for training Mixture of Gaussians, Bayes nets, HMMs, Combining labeled and unlabeled data: EM, reweighting labeled data, Co-training unlabeled data and model selection, Dimensionality reduction (PCA, SVD), Feature selection, HMMs: Forwards-Backwards, Viterbi, Supervised learning, Graphical Models: Representation, Inference, Learning, BIC			
Learning to make decisions: Markov decision processes, Reinforcement learning			

Additional Materials:

Basic probability and statistics. lecture 1, lecture 2, notes

Multivariate Gaussians. lecture