

machinelearning201: FrontPage

Machine Learning 201

Instructor: Dr. Michael Bowles & Dr. Patricia Hoffman

Overview of the Course

Machine Learning 201 and 202 cover topics in greater depth than 101 and 102. Participants in the class should come away able to read the current literature and apply what they read to their own work. Machine Learning 201 and 202 can be taken in any order.

Machine Learning 201 begins with ordinary least squares regression and extends this basic tool in a number of directions. We'll consider various regularization approaches. We'll introduce logistic regression and we'll learn how to code categorical inputs and outputs. We'll look at feature space expansions. These will lead naturally to generalizations of linear regression, known as the "generalized linear model" and the "generalized additive model".

Text: "The Elements of Statistical Learning - Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman

See also Prof Robert Tibshirani's notes for stats 315a: <http://www-stat.stanford.edu/~tibs/stat315a.html>

Prerequisites

Machine Learning 201 and 202 employ beginner-level probability, calculus and linear algebra (e.g. peruse the appendices in "Introduction to Data Mining" by Tan et. al. or [Linear Algebra](#), and [Probability Theory](#).) If you have taken Machine Learning [101](#) and [102](#) classes, you are well prepared for this course, but those are not required to start 201.

Participants should be familiar with R or be willing to pick R up outside of class. We will hand out R-code for most of our examples, but we won't spend time in 201 going through introductory material on R. Come to the first class with R loaded on your computer. <http://cran.r-project.org/> For your review, R are here: [References for R](#), [Reference for R Comments](#), [More R references](#). To integrate R with Eclipse [click here](#).

To get the most out of the class, participants will need to work through the homework assignments.

General Sequence of Classes:

Machine Learning 101: Supervised learning

Text: "Introduction to Data Mining", by Pang-Ning Tan, Michael Steinbach and Vipin Kumar

Machine Learning 102: Unsupervised Learning and Fault Detection

Text: "Introduction to Data Mining", by Pang-Ning Tan, Michael Steinbach and Vipin Kumar

Machine Learning 201: Advanced Regression Techniques, Generalized Linear Models, and Generalized Additive Models

Text: "The Elements of Statistical Learning - Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman

Machine Learning 202: Collaborative Filtering, Bayesian Belief Networks, and Advanced Trees

Text: "The Elements of Statistical Learning - Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman





Machine Learning Big Data: Adaptation and execution of machine learning algorithms in the map reduce framework

Machine Learning Text Processing: Machine learning applied to natural language text documents using statistical algorithms including indexing, automatic classification (e.g. spam filtering) part of speech identification, topic and modeling, sentiment extraction

Future Topics

Data Mining Social Networks
Text Mining
Recommender Methods
Big Data

Machine Learning 201 Syllabus:

<u>Week</u>	<u>Topics</u>	<u>Homework</u>	<u>Links</u>
1st Week	Advanced Regression Topics		Lecture 1 and 2
6/1/2011	Ordinary Least Squares - error bounds		
	Subset Select, fwd & backward step-wise		
	Least Angle Regression - LARS		
	Attribute basis change		
6/2/2011	Coefficient shrinkage methods	 Homework01.pdf	
	L1, L2 coefficient penalties		
	Ridge, lasso and elastic net		
2nd Week	Regression Topics		Lecture 3 and 4
6/8/2011	Logistic Regression	HW #1 Due	
6/9/2011	Attribute Expansion	 Homework02.pdf	
3rd Week	Factor Inputs/Outputs		Notes Week3
6/15/2011	Coding for Factor Inputs	HW #2 Due	
6/16/2011	Error-correcting codes	 Homework03.pdf	
4th Week	Generalized Linear Models		Notes Week4
6/22/2011		HW #3 Due	
6/23/2011		 Homework04.pdf	
5th	Paper on glmnet	http://www.jstatsoft.org/v33/i01/paper	Notes Week5

Week			
6/29/2011		HW #4 Due	
6/30/2011		Homework5	

We will be using the following text as a reference for the 201 and 202

"The Elements of Statistical Learning - Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. This is an excellent book. Virtually everyone in the field knows it and uses it as a standard reference. This book is free to look at on line. <http://www-stat.stanford.edu/~tibs/ElemStatLearn>

Anyone can read this web site, however only the instructors have permission to edit the site.