STAT 535: Foundations of Machine Learning	Fall 2017

Lecture 0: Syllabus

Lecturer: Fang Han — *August 06*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Lecturer.*

What will the course be about?

This is a 10-week course focused on introducing foundations of machine learning from philosophical, methodological, and theoretical perspectives. Physically, this is a course fully focused on "supervised learning", or even more narrowly, about "classification". Metaphysically, stemmed from this seemingly simple task, our ultimate goal is to appreciate and celebrate the statistical thinking.

Course overview

Task 1: (1st-2nd week, "as a philosopher") A unified view on supervised learning: risk minimization, PAC learning model, simple PAC learning theory, and the fundamental struggle (bias v.s. variance tradeoff).

Task 2: (2nd-3rd week, "as a methodologist") One thousand classifiers: Least squares regression, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), logistic regression, the perceptron algorithm, classification and regression tree (CART), random forest, the nearest-neighbor, the naive Bayes classifier, multi-layer neural network (deep learning), kernel trick, support vector machine (SVM), boosting, ...

Task 3: (4th-7th week, "as a mathematician") Statistical learning theory: growth function, VC dimension, generalization bound. A line of great minds (Chernoff, Hoeffding, Bennet, Bernstein, McDiarmid, Talagrand, Massart, Vapnik, Chervonenkis).

Task 4: (8th-10th week, "as a statistician") Understanding the classifiers: perceptron, SVM, reproducing kernel Hilbert space (kernel trick), and boosting.

References

The course is not built on any specific textbook, but an extraction and combination of the following ones:

1. Buhlmann and van der Geer (2011), Statistics for High-Dimensional Data: Methods, Theory and Applications;

2. Devroye, Gyorfi, and Lugosi (1997), A Probabilistic Theory of Pattern Recognition;

3. Ledoux and Talagrand (2011), Probability in Banach Spaces: Isoperimetry and Processes;

4. Pollard (1990), Empirical Processes: Theory and Applications;

5. Tsybakov (2008), Introduction to Nonparametric Estimation;

and publicly available lecture notes from Peter Bartlett, John Duchi, Sham Kakade, Marina Meila (the previous lecture notes on STAT535 are must-read!), Rob Schapire, Martin Wainwright, and Jian Zhang.

Prerequisites

This course is appropriate for a graduate student of a probability/statistics/machine learning background, and requires a certain (basic) level of mathematical maturity. A suggested list of course works (slightly revised based on Marina's):

1. A course in probability, including (i) weak and strong law of large numbers; (ii) basic and Lindberg-type central limit theorem; (ii) basics of multivariate analysis (conditional probability, independence, marginals, expectation, variance in multivariate setting);
2. Fundamentals of statistics: (i) basic distribution families; (ii) maximum likelihood estimation; (iii) estimating parameters of usual distributions (normal, multinomial);
3. Calculus and linear algebra: partial derivatives, gradient, the chain rule, vectors and matrices, matrix multiplications, eigenvalues and eigenvectors, positive definite matrices;
4. Algorithms and data structure at a basic level (arrays, lists, sets, $O(\cdot)$ notation).

Please do not hesitate to approach the instructor if you have any concern.

Evaluation

4 homework assignments (50%), a midterm exam (20%), and a final project (30%).

Miscellanea

Instructor: Fang Han (fanghan@uw.edu)

TA: Check the website

Lectures: Check the website

Office hour: Check the website