Large-scale Optimization and Learning:
A two-course sequence
Developing Tools For BIGDATA

Instructors: Constantine Caramanis and Sujay Sanghavi
Target Audience: Students from: ECE, CS, Math, ICES, DSSC
Year Offered: Fall 2012 – Spring 2013 (with goal of repeating alternate years)
Time: Tuesday & Thursday, 12:30 – 2:00 pm.

Course Design: The goal of this sequence is to teach a graduate level course, targeted to a broad audience of CommNetS students, general ECE students with strong technical background, as well as students from Math, CS, OR, Statistics, ICES, etc.

While we have designed the course as a two-semester sequence, it will be taught in such a way that students with an optimization background can directly take the second course, and likewise, students who only are interested in convex optimization and algorithms, can take only the first course.

Prerequisites: Students taking the Optimization course should have a relatively strong background in linear algebra. No prior knowledge of optimization is required. The Learning course (Part II of the sequence) requires additional background in probability and stochastic processes. Students without EE381J or equivalent should speak with the instructors. While no specific background in Machine Learning will be assumed, a course such as EE380L (Data mining) would be very helpful, and is recommended.

Motivation and Background: The last few years have seen tremendous attention in research, various sectors of industry (including startups), the media, etc., to problems in machine learning, in particular, problems in BIGDATA – including problems exhibiting an inherent high dimensionality, and problems involving truly massive data sets.

New ideas are required on two fronts: On the probability / statistics side, we have to rethink what kind of patterns we can find, what consistency and statistical reliability mean, what resilience to noise and corruption entails, etc., in such settings. On the algorithmic side, such massive problems require different computational approaches. In particular, interior point methods, SDP solvers, etc., that were the golden child of optimization in the last two decades, no longer seem appropriate in this new important context.

The philosophy and thus aim of both parts of this course will be to accomplish a judicious mix of theory and application: we will develop results in a rigorous way, proving key steps along the way, but homeworks will have a heavy computational element, featuring a series of what amounts to computational projects, requiring working with non-toy-size problems in Matlab.

The first course in the sequence will focus on Convex Optimization including basic material from convex geometry, convex analysis and convex optimization. It will cover basic modeling, and understanding how to find and exploit convexity. The second half of the course will develop algorithms for solving large scale optimization problems, particularly problems arising in large-scale machine learning problems. This includes gradient descent algorithms, stochastic gradient descent, proximal methods, accelerated descent algorithms, and so on.

The second course in the sequence will focus on Machine Learning. Conceptually, this is broken into three parts. The first part, introduces some of the classical problems, including classification and regression, and introduces some of the basic methods and ideas. The second part will focus on the high-dimensional regime. This addresses the important setting where the dimensionality of the estimation space or the sample space, may even exceed (possibly greatly exceed) the number of samples available. The third part will focus on massive data sets, and theoretical and algorithmic tools needed there. This will include, for instance, understanding what can be treated using MapReduce. Classroom instruction will focus on theoretical aspects of problems and solution approaches.

Topics Covered:

Part I: Convex Optimization

1. Background and Geometry. (8 lectures)
    a. Advanced Linear Algebra
    b. Convex Sets and Convex Functions
    c. Duality
2. Convex Optimization Problems and Applications: Recognizing and Exploiting Convexity. (8 lectures)
3. Algorithms. (8 lectures)
    a. Subgradient optimization and variants: Incremental subgradient, stochastic subgradient, etc.
    b. Augmented Lagrangian and Primal/Dual Methods.
    c. Proximal methods.

Course Text and Sources: **Convex Optimization** by S. Boyd and L. Vandenberghe, and various other sources.

Part II: Large-scale Learning

1. Machine Learning background. (6 Lectures)
    a. classification, regression, clustering, nearest neighbors, EM algorithm
    b. Spectral methods – SVD, PCA, Spectral clustering, MDS etc.
    c. Graphical models: how to model problems, Gibbs/MCMC sampling and inference, belief propagation.
2. High-dimensional learning. (12 lectures)
    a. The basics: Regression in high dimensions, concentration of measure (e.g., Johnson Lindenstrauss lemma, matrix concentration, etc.).
    b. Sparsity and model selection: representation in over-complete bases, compressed sensing, block/group sparsity, etc.
    c. Low-rank models: matrix completion, robust PCA etc., applications like collaborative filtering
    d. Low-complexity methods for sparse and low-rank models: greedy methods, iterative methods etc.
3. Big learning. (6-8 Lectures)

a. Divide and conquer (a.k.a. MapReduce) for machine learning: k-NN, SVM, PCA.
b. Subsampling and reduction strategies, bootstrapping, etc.
c. Convex optimization algorithms for massive data sets: gradient and stochastic gradient descent for SVM

Course Text and Sources: **TBA**