IE 598 – Big Data Optimization

Fall 2016

TR: 11:00AM – 12:20PM, 206 TB

Course Info

Instructor: **Niao He**
Email: niaohe@illinois.edu
Office: Transportation Building 211
Office Hour: Thursday, 10-11AM or by appointment via email

Course Description

Nearly every problem in machine learning and high-dimensional statistics can be formulated in terms of optimization of some function, possibly under some constraints. In the wake of Big Data era, there is an increasingly strong need to design efficient and scalable algorithms to address optimization problems in large scale - including problems exhibiting inherently high dimensions and problems involving truly massive data sets. The key aim of this course is to expose students to a wide range of modern algorithmic developments in optimization (mainly in convex optimization) and bring them near the frontier of research in large-scale optimization and machine learning.

Courtesy warning: The course will be theory-oriented and emphasize deep understanding of structures of optimization problems and computation complexity of numerical algorithms. The course will not cover any software or tools used for big data analytics and is not an application course.

Prerequisites: Students are expected to have strong knowledge of linear algebra, real analysis, and probability theory. Some prior exposure to optimization at a graduate level is preferred.

Textbooks: No book is required, but you are highly recommended to read the following:

- Ben-Tal & Nemirovski. *Lectures on Modern Convex Optimization*, SIAM. 2011
- Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer-Academic. 2003
- Boyd & Vandenberghe. *Convex Optimization*. Cambridge University Press. 2003
- Sra, Nowozin, Wright (eds). *Optimization for Machine Learning*. MIT Press. 2011
- Bubeck. *Convex Optimization: Algorithms and Complexity*. In Foundations and Trends in Machine Learning. 2015

Grading Policy: There will be no homework or exams. Grades will be based on

- *Scribing* (25%): each student will be assigned to scribe lectures one to two times.
- *Project* (50% report +25% presentation): This could be writing a survey on a certain topic based on several papers, conducting a novel large-scale experiment, or thinking about a concrete open theoretical question, applying optimization techniques to your own field, formalizing an interesting new topic, or trying to relate several problems. The end result should be a 10-15 page report, and a 25-30 minute presentation. I will provide a few ideas for possible topics for projects. You might also want to take a look at recent COLT, ICML, or NIPS proceedings.

Topical Outline and Tentative Schedule

- **Module I: Introduction and Fundamentals** (3 weeks)
  - Learning under the lenses of optimization
  - Basics of convex analysis
  - Conic programming: LP, SOCP, SDP
  - Structured optimization: convexity, smoothness, separability, etc.

- **Module II: Smooth Convex Optimization** (4 weeks)
  - Gradient descent method
  - Accelerated gradient descent methods (Nesterov's method and its variants)
  - Projection-free methods (Frank-Wolfe)
  - Coordinate descent methods
  - Applications in logistic regression, matrix completion, image processing, etc.

- **Module III: Nonsmooth Convex Optimizaton** (4 weeks)
  - Subgradient descent/Mirror descent
  - Smoothing techniques
  - Mirror-Prox method
  - Primal-Dual algorithms
  - Proximal algorithms
  - Applications in support vector machines, sparsity learning, etc.

- **Module IV: Stochastic and Online Optimization** (3 weeks)
  - Sample average approximation
  - Stochastic approximation (stochastic gradient descent)
  - Incremental gradient methods
  - Online algorithms
  - Applications in empirical risk minimization, online learning, etc.

- **Module V: Optimization Perspectives for Selective Topics in Machine Learning**
  - Sparsity learning
  - Large-scale kernel machines
  - Reinforcement learning
  - Deep learning