Large-scale Machine Learning and Optimization

Instructor: Dimitris Papailiopoulos, Assistant Professor, ECE Department
Email: dimitris@papail.io
Course website: papail.io/901
Class meets: Tuesday and Thursday, 1:00-2:15pm.
Student hours: Please email the instructor to schedule a time that is mutually convenient.

Course Outline:

This course will explore the mathematical foundations of a rapidly evolving new field: large-scale machine learning and optimization. We will focus on recent texts in machine learning, optimization, and randomized algorithms, with the goal to understand the tradeoffs that are driving algorithmic design in this new discipline. These trade-offs will revolve around statistical accuracy, scalability, algorithmic complexity, and implementation.

Sample topics include:

1. Optimization and Learning
- Stochastic Methods for Convex and Nonconvex Settings
- Constrained and Projection-free Optimization
- Overfitting, Generalization, and Algorithmic Stability

2. Large-scale Learning and Systems
- System Tradeoffs and Platforms
- Asynchronous Stochastic Optimization
- Distributed Learning

3. Matrix Methods for ML and Semidefinite Optimization
- Matrix Tools for Machine Learning
- Dimensionality Reduction
- Semidefinite Optimization and Randomized Rounding

Student Evaluation

- Semester Project: 50%

Milestone 1: mid semester project proposal, and presentation (15%).

Milestone 2: end-semester report, and poster presentation (35%).

- Paper Presentations: 30% (approximately bi-weekly frequency)

- Scribe Notes: 15%

- Student Participation: 5%

Semester project: Groups of 2-3 students will have to work on an open problem that is relevant to the course, or relevant to their own research and the course. The *first component* of the project will consist of a written (at most 3-page) project proposal that will be due approximately by the middle of the semester, followed by a slides presentation. During the presentation (expected to last approximately 15 minutes) each group will receive feedback from the instructor and the rest of the class. The *second component* of the project will be an end-of-semester report, where the members of each group will present the progress of their project, their potential innovations, and their concluding remarks. All projects will be accompanied by a poster. All posters will be presented during a poster day (at a date to be determined), where several other faculty and students will be invited to attend.

Paper presentations: Starting approximately at week #4 of lectures, and with an approximately bi-weekly frequency, the students will form groups and present assigned research papers. The presentations will be using slides, and are expected to be 15 minutes long.

Scribing: All students are required to scribed notes for at least one lecture. The scribes will be due one week after their corresponding lecture. Depending on the size of the class, up to B students will be selected per lecture, so that #enrolled_students/B = #lectures.

Student Participation: The students are expected to actively participate in the class with questions and comments. Furthermore, during the two phases of their semester project, and with a bi-weekly frequency, each group is expected to give a 5 minute progress report during one of the two lectures of the week. The purpose of these 5 minute progress reports, is so that each group receives feedback from the rest of the class and the instructor.

Key References

[1] **"*Convex Optimization: Algorithms and Complexity*"**
by Sébastien Bubeck
Source: http://research.microsoft.com/en-us/um/people/sebubeck/book.html

[2] **"*Understanding Machine Learning: From Theory to Algorithms*"**
by Shai Ben-David and Shai Shalev-Shwartz
Source: http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/index.html

[3] **"*Stability and Generalization*"**
by Olivier Bousquet and André Elisseeff
Source: http://www.jmlr.org/papers/volume2/bousquet02a/bousquet02a.pdf

[4] **"*Optimization Methods for Large-Scale Machine Learning*"**
by Léon Bottou, Frank E. Curtis, and Jorge Nocedal
Source: https://arxiv.org/pdf/1606.04838v1.pdf

[5] **"*Hogwild!: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent*"**
by Feng Niu, Benjamin Recht, Christopher Ré, and Stephen J. Wright
Source: https://people.eecs.berkeley.edu/~brecht/papers/hogwildTR.pdf

[6] **"*Foundations of Data Science*"**
By Avrim Blum, John Hopcroft, and Ravindran Kannan
Source: http://www.cs.cornell.edu/jeh/bookMay2015.pdf

Note: All presented material, slides, and scribes, will be posted online, on the course's webpage.

Prerequisites

This course is ideal for advanced graduate students, who are interested in applying novel research concepts to their own research. Students are expected to be familiar with basic concepts in optimization and machine learning, and a solid background in linear algebra and probability. Enrolled students are required to have attended at least 1 course, from at least 2 of the following categories, or equivalent, pending the instructor's approval.

ECE/CS/ME 532: Theory and Applications of Pattern Recognition
CS/ISyE 524 and ECE 601: Introduction to Optimization

ECE 729 Theory of Information Processing and Transmission
ECE 730: Probability and Stochastic Processes

ECE 901: Statistical Learning Theory
ECE 830: Statistical signal processing
CS 761: Advanced Machine Learning