Graduate Topic Course - STOR 893

Optimization for Machine Learning and Data Analysis

(Fall 2020)

Course overview

This is a special topic course taught at the Department of Statistics and Operations Research, UNC-Chapel Hill. The primary goal is to discuss theory and recent methods for optimization problems used in machine learning (ML) and data analysis. The content of this course consists of 4 parts ranging from representative models and background to algorithms. It also covers convergence guarantees, complexity bounds, efficient implementation, and concrete numerical examples. The course will discuss both deterministic and stochastic algorithms such as accelerated gradient, operator splitting, and primal-dual methods, and stochastic schemes such as stochastic gradient, variance reduction, and coordinate descent methods. It covers both convex and nonconvex models.

The course is designed for graduate students who have some background in applied math such as linear algebra, multivariable analysis, and computational skills. Background in convex analysis, numerical linear algebra, algorithms, machine learning, or statistics is also preferable to better follow the course.

Note: This course is not a machine learning or data science course.

Time and Place

***Lectures*: Tuesdays and Thursdays, 1:15PM - 2:30PM.**
***Place*: Coker Hall, Room 0201.** (This course is now moved to online due to students' preference)
***Zoom's link*:** https://unc.zoom.us/j/99933790086?pwd=cTRxTVZIYktidTRPU0lpMERkWW5odz09
(Passcode: 122521).

Instructor

Instructor: Quoc Tran-Dinh (`quoctd@email.unc.edu`)
Personal webpage: `http://quoctd.web.unc.edu`.

Office: 333 Hanes Hall, UNC-Chapel Hill.

Course content

This course consists of four parts:

1. ***Background and Mathematical Tools***
2. ***Large-Scale Convex Optimization***
3. ***Minimax Problems and Constrained Convex Optimization***
4. ***Nonconvex Optimization.***

Depending on time and progress, some topics may be given for self-studying and some topics may have more emphasis. The four parts of this course will cover the following concrete topics.

Week 1: Optimization formulations and background in the context of ML and data analysis.
Week 2: Representative optimization models in ML and data analysis.
Week 3: Brief overview on convex analysis and mathematical tools
Week 4: Gradient descent and fast gradient descent methods (different variants).
Week 5: Proximal operators and [fast] proximal gradient methods.

Week 6: Stochastic gradient methods (different variants).
Week 7: Variance reduction methods (SAGA, SVRG, and SARAH, etc).
Week 8: Smoothing techniques in nonsmooth optimization
Week 9: Frank-Wolfe, coordinate descent, and mirror descent methods (if time permits).
Week 10: Minimax problems, primal-dual first-order and operator splitting methods
Week 11: Nonconvex optimization: theory, algorithms, and applications.
Week 12: DC (difference-of-two-convex functions) programming and variants, nonconvex proximal alternating minimization algorithms (PALM), and nonconvex alternating direction method of multipliers (ADMM). Representative examples.
Week 13: Some optimization models and methods in neural networks and deep learning.
Week 14: Summary and Final Projects (Backup and Part 1).
Week 15: Final Projects (Part 2).

In the first two weeks, we will discuss several mathematical optimization models used in data analysis and machine learning. We will start from concrete models such as support vector machine (SVM), linear least-squares, LASSO, logistic regression, Poisson imaging reconstruction, optimal transport, inverse covariance estimation, portfolio, and robust PCA to abstract ones such as empirical risk minimization, and neural network training models (if time permits). These models are well-known in ML, but we will also discuss their variants and some possible extensions.

Next, we will review some mathematical background related to optimization, complexity theory, machine learning, and possibly some tools in statistics and probability theory. Then, we will study different methods for solving different classes of problems we discussed above. Each method can be applicable or efficiently apply to solve some of these models. Therefore, we need to study different methods so that we know which one should be suitable to solve which problem. In each method, we will discuss some principles behind to derive it, assumptions for convergence guarantees, its complexity and/or convergence rates, implementation remarks, and some concrete applications.

Finally, since it is a special topic, students are encouraged to engage the materials into their research. At the end of this course, we will save two weeks for students to present their final project. The final project could be a concrete topic related to student's own research or could be some team project assigned by the instructor. More details of the final project will be discussed in class, but it is quite flexible for students to propose more thoughts.

Course materials

Lecture notes

Lecture notes and slides will be provided to students via Sakai. Note that the instructor has only sides and some part of lecture notes. They must be used internally in the course. Please do not distribute these materials without the instructor's permission since they are under development. Any feedback is highly appreciated.

Books

Here are some books which contain some parts of the lectures

[B_1]. R. T. Rockafellar: Convex Analysis, 1970, Princeton Univ. Press. This book is now available online for free and can be downloaded from http://www.convexoptimization.com/TOOLS/ConvexAnalysisRockafellar.pdf.

[B_2]. S. Boyd and L. Vandenberghe: Convex Optimization, 2006, Cambridge Univ. Press. This book is available for free at http://stanford.edu/~boyd/cvxbook/.

[B_3]. Y. Nesterov: Introductory lectures on Convex Optimization, 2004 or 2018. The lectures can be found at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.855&rep=rep1&type=pdf.

[B_4]. H. H. Bauschke and P. Combettes: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer-Verlag, 2017.
[B_5]. D. Bertsekas, Convex Optimization Theory/Algorithms, Athena Scientific, 2009.
[B_6]. J. Nocedal and S. Wright, Numerical Optimization, Springer-Verlag, 2006.
[B_7]. G. Lan, First-order and Stochastic Optimization Methods for Machine Learning, Springer-Nature, 2020.

Other materials

These are good surveys/lecture notes for the course
[B_1]. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, The MIT Press, 2016.
[S_1]. S. Boyd et al: Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning, 3(1):1-122, 2011.
[S_2]. N. Parikh and S. Boyd: Proximal algorithms, Foundations and Trends in Optimization, 1(3):123-231, 2014.
[S_3]. S. Wright, Optimization Algorithms in Data Analysis. This survey paper is available online at http://www.optimization-online.org/DB_FILE/2016/12/5748.pdf
[L_1]. S. Bubeck, Convex Optimization: Algorithms and Complexity. This lecture note can be downloaded from http://arxiv.org/abs/1405.4980.
[L_2]. Prateek Jain and Purushottam Kar. Non-convex Optimization for Machine Learning. Foundations and Trends in Machine Learning, 10(3-4): 142-336, 2017.

Course Grade

- **Homework assignments:** A few homework assignments will be given during class. They will be counted for 30% of the final grade.
- **Take-home midterm:** A short take-home midterm exam will be given. It will be counted for 20% of the final grade.
- **Course projects:** Students work on projects (in teams or individually). They will be counted for 50% of the final grade. Students can select one of the following two formats:
  - Students are asked to read one or a few papers, or book chapters, then write a short report (between 4 and 8 pages) and present it in class. Some implementation and concrete applications outside such reading works are highly recommended.
  - Students are asked to work on an optimization problem, and implement some algorithms to solve it, then test the algorithms on synthetic and/or real datasets. Finally, they need to write a short report (between 4 and 8 pages) and present it in class. It can be some research problems, for which students are willing to share and discuss with their peers.

  Each student can work in team (often two to four students per team) or on her/his own project, which may relate to her/his research. Note that the optimization model in the final project should be nontrival, not just a simple exercise or homework, e.g., from other courses.
- **Exams:** There will be no written exam.

Community Standards in Our Course and Mask Use

This fall semester, while we are in the midst of a global pandemic, all enrolled students are required to wear a mask covering your mouth and nose at all times in our classroom. This requirement is to protect our educational community – your classmates and me – as we learn together. If you choose not to wear a mask, or wear it improperly, I will ask you to leave immediately, and I will submit a report to the Office of Student Conduct. At that point you will be disenrolled from this course for the protection of our educational community. An exemption to the mask wearing community standard will not typically be considered to be a reasonable accommodation. Individuals with a disability or health condition that prevents them from safely wearing a face mask must seek alternative accommodations through the Accessibility Resources and Service. For additional information, see Carolina Together.