Large-scale Optimization for Machine Learning

Aryan Mokhtari, PhD
Postdoctoral Associate
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Abstract: In large-scale data science, we train models for datasets containing massive numbers of samples. Training is often formulated as the solution of empirical risk minimization (ERM) problems which are optimization programs whose complexity scales with the number of elements in the dataset. This motivates the use of stochastic optimization techniques which, alas, come with their own set of limitations. In this talk, we will discuss recent developments to accelerate the convergence of stochastic optimization through the exploitation of second-order information. In particular, we present stochastic variants of quasi-Newton methods which approximate the curvature of the objective function using stochastic gradient information. We will explain how this leads to faster convergence and introduce an incremental method that exploits memory to achieve a superlinear convergence rate. This is the best-known convergence rate for a stochastic optimization method. We will also cover adaptive sample size schemes which rethink ERM as a collection of nested ERM problems in which the dataset grows at a geometric rate -- as opposed to stochastic methods in which samples are processed sequentially. We show how second-order versions of adaptive sample size methods are guaranteed to solve ERM problems to their statistical accuracy in just two passes over the dataset. We further extend this idea to the nonconvex setting to come up with computationally efficient methods for finding a local minimizer of ERM problems when the population risk is strongly Morse.

Bio: Aryan Mokhtari is a Postdoctoral Associate in the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT), since January 2018. Before joining MIT, he was a Research Fellow at the Simons Institute for the Theory of Computing at the University of California, Berkeley, for the program on "Bridging Continuous and Discrete Optimization", from August to December 2017. Prior to that, he was a graduate student at the University of Pennsylvania (Penn) where he received his M.Sc. and Ph.D. degrees in electrical and systems engineering in 2014 and 2017, respectively, and his A.M. degree in statistics from the Wharton School in 2017. Dr. Mokhtari received his B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2011. His research interests include the areas of optimization, machine learning, and artificial intelligence. His current research focuses on the theory and applications of convex and non-convex optimization in large-scale machine learning and data science problems. He has received a number of awards and fellowships, including Penn's Joseph and Rosaline Wolf Award for Best Doctoral Dissertation in electrical and systems engineering and the Simons-Berkeley Fellowship.