Beyond stochastic gradient descent for large-scale machine learning

Francis Bach, INRIA - ENS

2016 June, Thursday 07th

Many machine learning and statistics problems are traditionally cast as convex optimization problems. A common difficulty in solving these problems is the size of the data, where there are many observations ("large n") and each of these is large ("large p"). In this setting, online algorithms such as stochastic gradient descent which pass over the data only once, are usually preferred over batch algorithms, which require multiple passes over the data. Given n observations/iterations, the optimal convergence rates of these algorithms are $O(1/\sqrt{n})$ for general convex functions and reaches $O(1/n)$ for strongly-convex functions. In this talk, I will show how the smoothness of loss functions may be used to design novel simple algorithms with improved behavior, both in theory and practice : in the ideal infinite-data setting, an efficient novel Newton-basedstochastic approximation algorithm leads to a convergence rate of $O(1/n)$ without strong convexity assumptions. (joint work with Alexandre Defossez, Aymeric Dieuleveut, Nicolas Flammarion, and Eric Moulines)

Séminaire de Mathématiques Appliquées et de Calcul Scientifique du CERMICS
École des Ponts ParisTech, 6-8 avenue Blaise-Pascal, 77455 Champs-sur-Marne, France
http://cermics.enpc.fr/seminaires/cs/index.html