Stochastic gradient descent

Pontus Giselsson

Lund University, Sweden

Felix Klein Autumn Workshop "Continuous Optimization", September 16-18, 2020 organized by the Fraunhofer Institute for Industrial Mathematics ITWM in Kaiserslautern, Germany

Abstract

Problems with finite sum structure, such as many training problems in machine learning, are often solved using stochastic gradient descent. Finite sum structure means that the objective to be minimized is a sum of many functions, e.g., one for each training example in machine learning. Stochastic gradient descent picks one of the summands at random and performs a gradient step w.r.t. that function.

In this lecture, we will analyze the stochastic gradient descent method in nonconvex and convex settings and show its sublinear convergence. We will base the analysis on a general framework for showing sublinear convergence using Lyapunov inequalities. We will also discuss common variations of stochastic gradient descent such as AdaGrad and Adam.