Abstract

In classical optimization, one needs to calculate a full (deterministic) gradient of the objective function at each step, which can be extremely costly for modern applications of big data machine learning. A remedy to this problem is to approximate each full gradient with a random sample over the data. This approach reduces the computational cost at each step, but introduces statistical variance.

In this talk, I will present some recent progresses on applying variance reduction techniques previously developed for statistical Monte Carlo methods to this new problem setting. The resulting stochastic optimization methods are highly effective for practical big data problems in machine learning, and the new methods have strong theoretical guarantees that significantly improve the computational lower bounds of classical optimization algorithms.

Collaborators: Rie Johnson, Shai Shalev-Schwartz, Jialei Wang