
COURTNEY PAQUETTE, McGill University

SGD in the Large: Average-case Analysis, Asymptotics, and Step-size Criticality

In this talk, I will present a framework, inspired by random matrix theory, for analyzing the dynamics of stochastic gradient descent (SGD) when both the number of samples and dimensions are large. Using this new framework, we show that the dynamics of SGD on a least squares problem with random data become deterministic in the large sample and dimensional limit. Furthermore, the limiting dynamics are governed by a Volterra integral equation. This model predicts that SGD undergoes a phase transition at an explicitly given critical stepsize that ultimately affects its convergence rate, which we also verify experimentally. Finally, when input data is isotropic, we provide explicit expressions for the dynamics and average-case convergence rates. These rates show significant improvement over the worst-case complexities.