AUKOSH JAGANNATH, University of Waterloo
Online SGD on non-convex losses from high-dimensional inference

Stochastic gradient descent (SGD) is a popular tool in data science. Here one produces an estimator of an unknown parameter from independent samples of data by iteratively optimizing a loss function, which is random and often non-convex. We study the performance of SGD from an uninformative (random) start in the setting where the parameter space is high-dimensional. We develop nearly sharp thresholds for the number of samples needed for consistent estimation as one varies the dimension. They depend only on an intrinsic property of the population loss, called the information exponent and do not assume uniform control on the loss itself (e.g., convexity or Lipschitz-type bounds). These thresholds are polynomial in the dimension and the precise exponent depends explicitly on the information exponent. As a consequence, we find that except for the simplest tasks, almost all of the data is used in the initial search phase, i.e., just to get non-trivial correlation with the ground truth, and that after this phase, the descent is rapid and exhibits a law of large numbers. We illustrate our approach by applying it to a wide set of inference tasks such as parameter estimation for generalized linear models and spiked tensor models, phase retrieval, online PCA, as well as supervised learning for single-layer networks with general activation functions. Joint work with G. Ben Arous (NYU Courant) and R. Gheissari (Berkeley)