MAHDI SOLTANOLKOTABI, University of Southern California

From Shallow to Deep: Rigorous Guarantees for Training Neural Networks

Neural network architectures (a.k.a. deep learning) have recently emerged as powerful tools for automatic knowledge extraction from data, leading to major breakthroughs in a multitude of applications. Despite their wide empirical use the mathematical success of these architectures remains a mystery. A major challenge is that training neural networks correspond to extremely high-dimensional and nonconvex optimization problems and it is not clear how to provably solve them to global optimality. While training neural networks is known to be intractable in general, simple local search heuristics are often surprisingly effective at finding global/high quality optima on real or randomly generated data. In this talk I will discuss some results explaining the success of these heuristics. First, I will discuss results characterizing the training landscape of single hidden layer networks demonstrating that when the number of hidden units are sufficiently large then the optimization landscape has favorable properties that guarantees global convergence of (stochastic) gradient descent to a model with zero training error. Second, I introduce a de-biased variant of gradient descent called Centered Gradient Descent (CGD). I will show that unlike gradient descent, CGD enjoys fast convergence guarantees for arbitrarily deep convolutional neural networks with large stride lengths. The second part of the talk is joint work with Samet Oymak.