

University of Chicago and Toyota Technological Institute at Chicago
Machine Learning Seminar Series

PRESENTS:



Greg Ongie
University of Chicago



Blake Woodworth
TTIC

Greg Ongie

Title: A function space view of overparameterized neural networks

Abstract: Contrary to classical bias/variance tradeoffs, deep learning practitioners have observed that vastly overparameterized neural networks with the capacity to fit virtually any labels nevertheless generalize well when trained on real data. One possible explanation of this phenomenon is that complexity control is being achieved by implicitly or explicitly controlling the magnitude of the weights of the network. This raises the question: What functions are well-approximated by neural networks whose weights are bounded in norm? In this talk, I will give some partial answers to this question. In particular, I will give a precise characterization of the space of functions realizable as a two-layer (i.e., one hidden layer) neural network with ReLU activations having an unbounded number of units, but where the Euclidean norm of the weights in the network remains bounded. Surprisingly, this characterization is naturally posed in terms of the Radon transform as used in computational imaging, and I will show how tools from Radon transform analysis yield novel insights about learning with two and three-layer ReLU networks.

Blake Woodworth

Title: The complexity of finding stationary points in convex and non-convex optimization

Abstract: Non-convex optimization algorithms typically guarantee convergence to approximate stationary points of the objective. However, the fundamental complexity of finding such points is poorly understood, even in the convex setting, and especially in comparison to our very thorough understanding of the complexity of finding points with near optimal function value in the convex case. In this talk, I will discuss two recent papers in which we tightly bound the stochastic first-order oracle complexity of finding an approximate stationary point, first for the convex case and then for the non-convex case. An important implication of our work is that, in a certain sense, plain SGD is an optimal algorithm for stochastic non-convex optimization.

Host: Nati Srebro and Rebecca Willett

Friday, November 15, 10:30 – 11:30 am
JCL 390

(Remote broadcast at TTIC, 6045 S Kenwood Avenue, 5th Floor, Room 526)