THE UNIVERSITY OF CHICAGO

Computational and Applied Mathematics & Statistics Student Seminar

---

Suzanna Parkinson

Computational and Applied Mathematics
University of Chicago

Depth Separation of Neural Networks in Learning

Tuesday, February 13, 2024
12:30 PM
Searle 240A

ABSTRACT

It has been empirically observed that deeper neural networks often outperform shallow networks even though both models are universal approximations of continuous functions, but we do not yet fully understand why. We establish a separation between the learning capabilities of two-layer vs. three-layer feedforward ReLU neural networks. Although previous work has shown depth separation in terms of the approximation capabilities or expressivity of neural networks of varying depths, our result focuses on separation in terms of generalization performance and sample complexity. We consider neural networks trained using empirical risk minimization with weight decay regularization and polynomially many samples from a particular input distribution. We demonstrate that there is a family of label distributions that even arbitrarily wide two-layer networks trained with this procedure cannot learn, but narrow three-layer networks can. We also show that the reverse cannot occur; any labeling distributions that two-layer networks trained with this procedure can learn can also be learned by three-layer networks.