

CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday – June 14, 2019

Paper review :The Convergence Rate of Neural Networks for Learned Functions of Different Frequencies

Guofei Pang

We study the relationship between the speed at which a neural network learns a function and the frequency of the function. We build on recent results that show that the dynamics of overparameterized neural networks trained with gradient descent can be well approximated by a linear system. When normalized training data is uniformly distributed on a hypersphere, the eigenfunctions of this linear system are spherical harmonic functions. We derive the corresponding eigenvalues for each frequency after introducing a bias term in the model. This bias term had been omitted from the linear network model without significantly affecting previous theoretical results. However, we show theoretically and experimentally that a shallow neural network without bias cannot learn simple, low frequency functions with odd frequencies, in the limit of large amounts of data. Our results enable us to make specific predictions of the time it will take a network with bias to learn functions of varying frequency. These predictions match the behavior of real shallow and deep networks.