

Deep Learning meets statistics: Improving neural networks with statistical theory

Sophie Langer.
TU Darmstadt.

25-th June 2021, 1pm UK time

Abstract

In this talk we analyse the L2 error of neural network regression estimates with one hidden layer. Under the assumption that the Fourier transform of the regression function decays suitably fast, we show that an estimate, where all initial weights are chosen according to proper uniform distributions and where the weights are learned by gradient descent, achieves a rate of convergence of $1/n$ (up to a logarithmic factor). Our statistical analysis implies that the key aspect behind this result is the proper choice of the initial inner weights and the adjustment of the outer weights via gradient descent. This indicates that we can also simply use linear least squares to choose the outer weights. We prove a corresponding theoretical result and compare our new linear least squares neural network estimate with standard neural network estimates via simulated data. Our simulations show that our theoretical considerations lead to an estimate with an improved performance. Hence the development of statistical theory can indeed improve neural network estimates.