On the Adaptivity of Deep Neural Networks

Yu-Xiang WANG

Department of Computer Science, University of California, Santa Barbara

Email: yuxiangw@cs.ucsb.edu

Why do deep neural networks work better in practice than classical methods, e.g., kernels / splines? The talk covers a recent line of research that inspects DNNs from the classical non-parametric regression (or "curve fitting") which reveals that the reason why DNNs work better might be due to its adaptivity when we tune its standard hyperparameters, which implicitly discovers hidden sparsity and low-dimensional structures. The speaker will go over theory and examples to illustrate this point. For example, the effect of "Weight Decay" is in fact to select a small number of learned basis functions and the deeper the network is, the sparser it gets no matter how overparameterized the network is. Also, weight-decay helps a ConvResNeXT architecture to discover avoid the curse-of-dimensionality when the high-dimensional data is embedded in an unknown low-dimensional manifold. Last but not least, the speaker will explain how the choice of "learning rate" determines the smoothness of functions that SGD can stably converge to, thus informing us on generalization.

References:

[1] Deep Learning meets Nonparametric Regression: Are Weight-Decayed DNNs Locally Adaptive? Kaiqi Zhang, Yu-Xiang Wang. ICLR 2023.

[2] Nonparametric Classification on Low Dimensional Manifolds using Overparameterized Convolutional Residual Networks. Kaiqi Zhang, Zixuan Zhang, Minshuo Chen, Mengdi Wang, Tuo Zhao, Yu-Xiang Wang. Under review.

[3] Adaptivity by tuning learning Rate: Minima-stability and Generalization in the Noisy Case. Kaiqi Zhang, Daniel Soudry, Yu-Xiang Wang. In preparation.