Approximation theory of deep convolutional networks

DING-XUAN ZHOU

School of Data Science and Department of Mathematics, City University of Hong Kong, Hong Kong

Email: `mazhou@cityu.edu.hk`

Deep learning has been widely applied and brought breakthroughs in speech recognition, computer vision, and many other domains. The involved deep neural network architectures and computational issues have been well studied in machine learning. But there lacks a theoretical foundation for understanding the approximation or generalization ability of deep learning models with network architectures such as deep convolutional neural networks (CNNs) with convolutional structures. The convolutional architecture gives essential differences between the deep CNNs and fully-connected deep neural networks, and the classical approximation theory of fully-connected networks developed around 30 years ago does not apply. This talk describes an approximation theory of deep CNNs. We show the universality of a deep CNN, meaning that it can be used to approximate any continuous function to an arbitrary accuracy when the depth of the neural network is large enough. Rates of approximation are given in terms of the number of free parameters to be trained, verifies the efficiency of deep CNNs in dealing with large dimensional data. We also introduce a downsampled deep CNN and show that any fully-connected neural network can be realized by such a deep CNN, which verifies that the modelling ability of deep CNNs is at least as good as that of fully-connected networks. Examples are also given to demonstrate advantages of deep CNNs.