Extended abstract:

This course covers foundation and recent advances of deep neural networks (DNNs) from the point of view of statistical theory. Understanding the power of DNNs theoretically is arguably one of the greatest problems in machine learning. During the last decades DNNs have made rapid process in various machine learning tasks like image and speech recognition and game intelligence. Unfortunately, little is yet known about why this method is so successful in practical applications. Recently, there are different research topics to also prove the power of DNNs from a theoretical point of view. From an aspect of statistical theory, several results could already show good convergence results for DNNs in learning different function classes.

The course is roughly divided into three parts:

Part 1: Theory of shallow neural networks. After a quick introduction into DNNs and a discussion of different network architectures, we start our theoretical analysis with the universal approximation theorem. Here we discuss different proof strategies that give us insights into the approximation power of shallow neural networks. Later on we show convergence rates for estimators based on shallow neural networks. Finally, we compare shallow neural networks with deep ones and show advantages of multiple hidden layers.

Part 2: Theory of deep ReLU networks. In the second part we discuss estimation rates for deep ReLU networks. Therefore we analyze the approximation properties of ReLU networks in terms of network depth and width. Especially their efficient approximation of smooth functions is of interest. Based on this we obtain risk bounds for fully connected networks. Here we discuss different settings where neural network estimators are able to circumvent the curse of dimensionality.

Part 3: Theory for networks learned by gradient descent. In our third part we discuss a rate of convergence results for networks that are trained by gradient descent. Here we consider shallow networks and a special class of regression functions. As an outlook we formulate possible extension of our result and discuss important steps to develop an encompassing theory of deep learning.