CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday - June 26, 2020

Neural Tangent Kernel: Convergence and Generalization in Neural Networks

Guofei Pang

At initialization, artificial neural networks (ANNs) are equivalent to Gaussian processes in the infinite-width limit, thus connecting them to kernel methods. We prove that the evolution of an ANN during training can also be described by a kernel: during gradient descent on the parameters of an ANN, the network function (which maps input vectors to output vectors) follows the so-called kernel gradient associated with a new object, which we call the Neural Tangent Kernel (NTK). This kernel is central to describe the generalization features of ANNs. While the NTK is random at initialization and varies during training, in the infinite-width limit it converges to an explicit limiting kernel and stays constant during training. This makes it possible to study the training of ANNs in function space instead of parameter space. Convergence of the training can then be related to the positive-definiteness of the limiting NTK. We then focus on the setting of least-squares regression and show that in the infinite-width limit, the network function follows a linear differential equation during training. The convergence is fastest along the largest kernel principal components of the input data with respect to the NTK, hence suggesting a theoretical motivation for early stopping. Finally we study the NTK numerically, observe its behavior for wide networks, and compare it to the infinite-width limit.