Approximation of Functions by Composition

Zuowei Shen
Department of Mathematics, National University of Singapore, Singapore
Email: scidean@nus.edu.sg

The primary task in supervised learning is approximating/estimating a function $f$ through samples drawn from a probability distribution on the input space. Learning is the process of approximating $f$ by a function $g$ with tunable parameters, which can be adjusted so that $g$ becomes close to $f$ in some averaged sense with respect to the input distribution. Usually, we pick a nice $g$ to work with. For regression problems, the simplest $g$ one can consider is an affine function, whose parameters can be fitted. For classification problems, one can consider $g$ an affine function followed by a sigmoid transformation and a maximization across output coordinates. This is known as logistic regression. Although such simple $g$’s are easy to analyze and optimize in practice, when the underlying $f$ is complex, they tend to have low approximation quality. The key idea in deep learning is to expand a simple approximator $g$ by composing with it a series of nested feature extractors, i.e. one finds $T_1, T_2, \ldots, T_N$ such that $g$ composes with $T_1, T_2, \ldots, T_N$ that approximates $f$. In this talk, we discuss mathematical theory behind such approximations and how the theory can be used to understand and design deep learning network; and how it differs from the classic approximation theory.