Benefit of deep learning: Efficiency of function estimation and its optimization guarantee

Taiji Suzuki 1,2

1) Department of Mathematical Informatics, The University of Tokyo, Tokyo, Japan
2) Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan

Corresponding Author: Taiji Suzuki, taiji@mist.i.u-tokyo.ac.jp

ABSTRACT

In this talk, I discuss how deep learning can statistically outperform shallow learning methods such as kernel methods from the viewpoint of statistical estimation. First, I will discuss the excess risk bounds of deep learning in the Besov space and its variants, in which sparsity and non-convex geometry of the target function space play the essential role. In particular, it is shown that deep learning can work for high-dimensional input while the linear estimators including kernel ridge regression suffers from curse of dimensionality. In the latter half, I present a deep learning optimization framework based on a noisy gradient descent in an infinite dimensional Hilbert space (gradient Langevin dynamics), and show generalization error and excess risk bounds for the solution obtained by the optimization procedure. The proposed framework can deal with finite and infinite width networks simultaneously unlike existing one such as neural tangent kernel and mean field analysis. Moreover, I will show that deep learning can avoid the curse of dimensionality in a teacher-student setting, and eventually achieve better excess risk than kernel methods.