



Optimization and Machine Learning Seminar

Optimal Rates for Generalization of Gradient Descent Methods with Deep Neural Networks

Dr. Junyu Zhou

Catholic University of Eichstatt-Ingolstadt, Germany

Abstract

Recent progress has been made in understanding the statistical generalization performance of gradient descent methods for overparameterized neural networks within the neural tangent kernel (NTK) regime. However, most of the existing work on regression problems is limited to shallow network architectures, leaving a notable gap in the theory for deep neural networks. We address this gap by presenting a comprehensive generalization analysis for deep ReLU networks trained using gradient descent (GD) and stochastic gradient descent (SGD). Specifically, we establish the first known minimax-optimal rates of excess population risk for both GD and SGD with deep ReLU networks, under the assumption that the network width scales polynomially with respect to the network depth and training sample size. Our results demonstrate that with sufficient width, gradient descent methods for deep ReLU networks can achieve the optimal rates of generalization on par with kernel methods.

Date: March 9, 2026 (Monday)

Time: 4:00 pm - 5:00 pm

Venue: Room 210, Run Run Shaw
Building HKU