復旦大學
FUDAN UNIVERSITY

Applied Math Ph.D. Seminar

Phenomenon-Driven Deep Learning Theory: From Implicit Regularization in Matrix Factorization to Loss Spike Mechanisms in Adam

Speaker: Zhiwei Bai (Shanghai Jiao Tong University)

Time: 2025-10-16, 16:10 to 17:00

Location: Rm 1801, Guanghua East Tower

Advisor: Yaoyu Zhang (Shanghai Jiao Tong University),
Zhiqin Xu (Shanghai Jiao Tong University)

Abstract: Deep neural networks, as highly nonlinear complex systems, present formidable theoretical challenges. Phenomenon-driven research—grounded in meticulous observation and carefully designed experiments to discover intrinsic system patterns—offers a crucial gateway to understanding these complex systems. This talk presents our recent advances in deep learning generalization and optimization theory through a phenomenon-driven approach.

One of the most counterintuitive phenomena in modern machine learning is that neural networks maintain excellent generalization despite overparameterization. Understanding implicit regularization mechanisms in overparameterized models has become essential to deep learning theory. Matrix factorization models, as an important subclass, provide an ideal testbed for studying implicit regularization. This talk first reviews the generalization puzzle, and introduces our discovery of a fundamental structural property of loss landscapes: the Embedding Principle, which reveals an elegant inheritance relationship between critical points across networks of different scales. Building on this, we analyze matrix factorization training dynamics from a model-data decoupling perspective, elucidating when, how, and why different implicit regularization effects (low rank, low nuclear norm) emerge, providing a unified understanding of this system.

This talk also presents another phenomenon-driven study: loss spike—a sudden and sharp surge in the loss function that subsequently subsides. These spikes are observed across a wide range of network architectures and datasets, yet their underlying mechanisms remain elusive. While previous studies attributed loss spikes to complex loss landscape geometry, we find they originate from Adam's adaptive preconditioning mechanism. Specifically, when gradients in certain layers gradually diminish during training, the adaptive mechanism persistently pushes the maximum eigenvalue of the preconditioned Hessian above the stability threshold, triggering sustained instability. This result provides a novel theoretical perspective for understanding and controlling loss spike behavior.