



THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

**SEMINAR ON DATA SCIENCE
AND APPLIED MATHEMATICS**

**A Statistical View on Implicit Regularization:
Gradient Descent Dominates Ridge**

By

Dr. Jingfeng WU

UC Berkeley

Abstract

A key puzzle in deep learning is how simple gradient methods find generalizable solutions without explicit regularization. This talk discusses the implicit regularization of gradient descent (GD) through the lens of statistical dominance. Using linear regression as a clean proxy, we present three surprising findings.

First, GD dominates ridge regression: with comparable regularization, the excess risk of GD is always within a constant factor of ridge, but ridge can be polynomially worse even when tuned optimally. Second, GD is incomparable with online stochastic gradient descent (SGD). While it is known that for certain problems GD can be polynomially better than SGD, the reverse is also true: we construct problems, inspired by benign overfitting theory, where optimally stopped GD is polynomially worse. Finally, GD dominates SGD for a significant subclass of problems -- those with fast and continuously decaying covariance spectra -- which includes all problems satisfying the standard capacity condition.

This is joint work with Peter Bartlett, Sham Kakade, Jason Lee, and Bin Yu.

Bio: *Jingfeng Wu is a postdoctoral fellow at the Simons Institute for the Theory of Computing at UC Berkeley. His research focuses on deep learning theory, optimization, and statistical learning. He earned his Ph.D. in Computer Science from Johns Hopkins University in 2023. Prior to that, he received a B.S. in Mathematics (2016) and an M.S. in Applied Mathematics (2019), both from Peking University. In 2023, he was recognized as a Rising Star in Data Science by the University of Chicago and UC San Diego.*

Date : 11 October 2025 (Saturday)
Time : 10:30a.m.
Venue : Lecture Theatre F

All are welcome