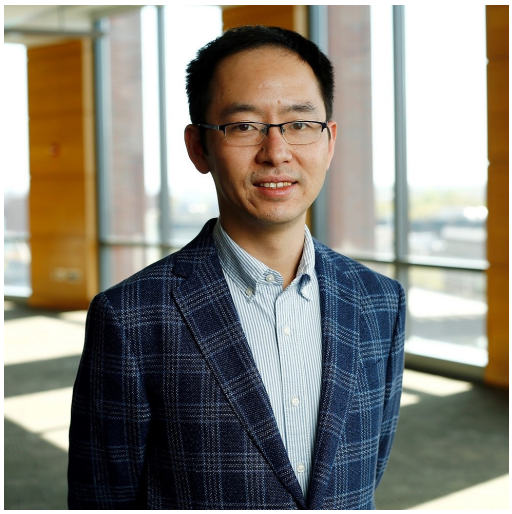
Department of Statistics Seminar Series



Yuxin Chen
Department of Statistics and Data Science
University of Pennsylvania

Transformers Meet In-Context Learning: A Universal Approximation Theory

Wednesday, March 11th, 2026
12:10-1:30 PM
Hill Center - 552

Zoom Meeting: Meeting ID: 926 1608 1995
Password: 153525

https://rutgers.zoom.us/j/92616081995?pwd=TqPVh6Y1O6aLUt8i2uUBCRhrgH6Ayx.1

Light refreshments will be served in Hill 452 11:40am

Abstract:

Large language models are capable of in-context learning, the ability to perform new tasks at test time using a handful of input-output examples, without parameter updates. We develop a universal approximation theory to elucidate how transformers enable in-context learning. For a general class of functions (each representing a distinct task), we demonstrate how to construct a transformer that, without any further weight updates, can predict based on a few noisy in-context examples with vanishingly small risk. Unlike prior work that frames transformers as approximators of optimization algorithms (e.g., gradient descent) for statistical learning tasks, we integrate Barron's universal function approximation theory with the algorithm approximator viewpoint. Our approach yields approximation guarantees that are not constrained by the effectiveness of the optimization algorithms being mimicked, extending far beyond convex problems like linear regression. The key is to show that (i) any target function can be nearly linearly represented, with small L1-norm, over a set of universal features, and (ii) a transformer can be constructed to find the linear representation -- akin to solving Lasso -- at test time. This is joint work with Gen Li, Yuchen Jiao, Yu Huang, and Yuting Wei.

Bio:

Yuxin Chen is currently a professor of statistics and data science and of electrical and systems engineering at the University of Pennsylvania. Before joining UPenn, he was an assistant professor of electrical and computer engineering at Princeton University. He completed his Ph.D. in Electrical Engineering at Stanford University and was also a postdoc scholar at Stanford Statistics. His current research interests include statistics, machine learning, and optimization. He has received the Alfred P. Sloan Research Fellowship, the Leo Breiman Junior Award, the SIAM Activity Group on Imaging Science Best Paper Prize, the ICCM Best Paper Award (gold medal), and has been selected as a finalist for the Best Paper Prize for Young Researchers in Continuous Optimization, and an Information Theory Society Distinguished Lecturer. He has also received the Princeton Graduate Mentoring Award.