THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

Statistics Colloquium

CHAO MA

Department of Mathematics
Stanford University

"Implicit bias of optimization algorithms and generalization of over-parameterized neural networks"

MONDAY, February 20, 2023 at 4:30 PM
Jones 303, 5747 S. Ellis Avenue
Refreshments will be served prior to the seminar in Jones 303 at 4:00 pm

ABSTRACT

Modern neural networks are usually over-parameterized—the number of parameters exceeds the number of training data. In this case the loss function tends to have many (or even infinite) global minima, which imposes a challenge of minima selection on optimization algorithms besides the convergence. Specifically, when training a neural network, the algorithm not only has to find a global minimum, but also needs to select minima with good generalization among many others. We study the mechanisms that facilitate global minima selection of optimization algorithms, as well as its connection with good generalization performance. First, with a linear stability theory, we show that stochastic gradient descent (SGD) favors global minima with flat and uniform landscape. Then, we build a theoretical connection of flatness and generalization performance based on a special multiplicative structure of neural networks. Connecting the two results, we develop generalization bounds for neural networks trained by SGD. We also study the behavior of optimization algorithms around manifold of minima and characterize their explorations from one minimum to another.