THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

Joint seminar with the Department of Computer Science

WEIHAO KONG

University of Washington

The Surprising Power of Little Data

WEDNESDAY, January 27, 2021 at 4:00 PM
Via Zoom (session information will be e-mailed to subscribers)

ABSTRACT

In this talk, I will discuss several examples of my research that reveal a surprising ability to extract accurate information from modest amounts of data.First, I'll discuss the problem of learning from a large number of heterogeneous data sources: A common modern machine learning scenario involves a large amount of data contributed by a large number of heterogeneous individuals, with each individual providing a modest amount of data. In such a scenario, it is crucial the model being trained for each individual leverages data from other individuals (who may have different underlying models). The fundamental question is: To what extent can a large number of data sources compensate for the lack of data from each source? We answer this question in two fundamental settings where each data source is 1) a binomial distribution, 2) a linear model, by providing optimal algorithms and matching lower bounds.

In the second part of the talk, I'll discuss estimating learnability: Without enough data to learn a good model for prediction, is it possible to tell whether a good model exists? Surprisingly, we showed this is possible under linear model assumptions in supervised learning and contextual bandits settings. I will also discuss the implications of our techniques to model selection in this contextual bandits/reinforcement-learning setting.