Statistics Colloquium

KATHRYN ROEDER

Department of Statistics
Carnegie Mellon University

"Model-free prediction test and distribution-free independence test for high dimension data with applications to genomics data"

MONDAY, NOVEMBER 7, 2022, at 4:30 PM
Jones 303, 5747 S. Ellis Avenue
Refreshments before the seminar at 4:00 PM in Jones 304.

ABSTRACT

Statistical theory has mostly focused on testing the dependence between covariates and response under parametric or semiparametric models. In reality, the model assumptions might be too restrictive to be satisfied, and it is of substantial interest to test the significance of the prediction in a completely model-free setting. Likewise, test of independence is of fundamental importance in modern data analysis, with broad applications in variable selection, graphical models, and causal inference. When the data are high dimensional and the potential dependence signal is sparse, independence testing becomes very challenging without distributional or structural assumptions. To solve both of these problems we utilize data splitting to allow us to borrow the strength of the most powerful machine learning regression algorithms and achieve solutions that are computationally efficient. Our proposed methods are nonparametric and can be applied to a wide range of real applications. We illustrate the ideas with genomics applications. This work is joint with Zhanrui Cai (Iowa State) and Jing Lei (CMU)