Data Science Seminar

Hosted by the Department of Mathematics and Statistics

- Date: Thursday, April 27, 2023
- Time: 1:15pm – 2:15pm
- Room: Whitney Hall 100E
- Speaker: Dr. Runze Li (Penn State University)
- Title: Model-Free Conditional Feature Screening with FDR Control.

Abstract

In this talk, I will present a model-free conditional feature screening method with false discovery rate (FDR) control for ultra-high dimensional data. The proposed method is built upon a new measure of conditional independence. Thus, the new method does not require a specific functional form of the regression function and is robust to heavy-tailed responses and predictors. The variables to be conditional on are allowed to be multivariate. The proposed method enjoys sure screening and ranking consistency properties under mild regularity conditions. To control the FDR, we apply the Reflection via Data Splitting method and prove its theoretical guarantee using martingale theory and empirical process techniques. Simulated examples and real data analysis show that the proposed method performs very well compared with existing works.

Biography of the speaker: Dr. Li is the Eberly Family Chair Professor in Statistics, The Pennsylvania State University. He served as Co-Editor of Annals of Statistics from 2013 to 2015. Runze Li is Fellow of the American Association for the Advancement of Science (AAAS), the American Statistical Association (ASA), and the Institute of Mathematical Statistics (IMS). His recent honors and awards also include the Distinguished Achievement Award of International Chinese Statistical Association, 2017, Faculty Research Recognition Awards for Outstanding Collaborative Research. College of Medicine, Penn State University in 2018 and Distinguished Mentoring Award, Eberly College of Science, Penn State University in 2023. His research interests include theory and methodology in variable selection, feature screening, robust statistics, nonparametric and semiparameteric regression. His interdisciplinary research aims to promote the better use of statistics in social behavioral research, neural science research and climate studies.