

Variable Selection for High-Dimensional Incomplete Data

Philip Leung-Ho Yu

**Department of Mathematics and Information Technology
The Education University of Hong Kong**

Regression analysis is often impacted by several challenges such as high dimensionality, severe multicollinearity, and missing data. These problems can obscure important relationships and result in biased conclusions. In this talk, we propose a new method that effectively addresses these challenges by combining data imputation and variable selection. Specifically, our approach incorporates a novel multiple imputation algorithm based on matrix completion (Multiple Accelerated Inexact Soft-Impute), a more stable and accurate randomized lasso method (Hybrid Random Lasso), and a consistent method to integrate a variable selection method with multiple imputation. The proposed method is applied to analyze the Asian American minority subgroup in the 2017 National Youth Risk Behavior Survey. We focus on studying key risk factors related to the intention for suicide among Asian Americans. Through simulations and real data analyses on various regression and classification settings, our method demonstrates improved accuracy, consistency, and efficiency in both variable selection and prediction. This is a joint work with Lixing Liang and Yipeng Zhuang.