The Hong Kong University of Science and Technology

Joint Statistics Seminar

Department of Information Systems, Business Statistics and Operations Management

and

Department of Mathematics

Semi-supervised Inference for Explained Variance in High-dimensional Linear Regression and Its Applications

by

Dr. Zijian Guo

Department of Statistics and Biostatistics

Rutgers University

Date: July 19, 2018 (Thursday)

Time: 3:00 pm – 4:00 pm

Venue: Room 5047 (LSK Business Building)

Abstract

We consider statistical inference for the explained variance under the high-dimensional linear model in the semi-supervised setting. A calibrated estimator, which efficiently integrates both labelled and unlabelled data, is proposed. It is shown that the estimator achieves the minimax optimal rate of convergence in the general semi-supervised framework. The optimality result characterizes how the unlabelled data affects the minimax optimal rate. Moreover, the limiting distribution for the proposed estimator is established and data-driven confidence intervals for the explained variance are constructed. We further develop a randomized calibration technique for statistical inference in the presence of weak signals and apply the obtained inference results to a range of important statistical problems, including signal detection and global testing, prediction accuracy evaluation, and confidence ball construction. The numerical performance of the proposed methodology is demonstrated in simulation studies and an analysis of estimating heritability for a yeast segregant data set with multiple traits.

All interested are welcome!

For details, please contact ISOM Department.