Mathematics, Statistics, and Computer Science   **@ UIC**

Special Colloquium

Promoting Similarity of Sparsity Structures in Integrative Analysis

Yuan Huang (Yale University)

Abstract: For data with high-dimensional covariates but small sample sizes, the analysis of a single dataset often suffers from a lack of power and poor reproducibility. The integrative analysis of multiple independent datasets provides an effective way of pooling information and outperforms single-dataset and several alternative multi-datasets methods. In this study, we consider penalized variable selection and estimation in integrative analysis. Advancing from the existing studies, we introduce a novel penalty to explicitly encourage the similarity of sparsity structures. This study is motivated by the practical consideration that under many scenarios, multiple datasets are expected to share common important covariates. Theoretically the proposed method has established selection and estimation consistency properties under the high dimensional settings. Numerically the proposed method has identification and estimation performance better than or comparable to the alternatives under a wide spectrum of simulation scenarios. In the analysis of three lung cancer datasets with gene expression measurements, the proposed method identifies genes with sound biological implications and satisfactory prediction performance.

Wednesday, January 18 at 3:00 PM in SEO 636