

STATISTICS SEMINAR

Shuangge Ma, Ph.D.
Yale School of Public Health

Thursday, February 25, 2016
Dickens Hall, Room 207, 4:00-5:00 pm
Refreshments: Dickens 108, 3:30-4:00 pm



Promoting Similarity of Sparsity Structures in Integrative Analysis with Penalization

For data with high-dimensional covariates but small sample sizes, the analysis of single datasets often generates unsatisfactory results. The integrative analysis of multiple independent datasets provides an effective way of pooling information and outperforms single-dataset and several alternative multi-datasets methods. Under many scenarios, multiple datasets are expected to share common important covariates, that is, the corresponding models have similarity in their sparsity structures. However, the existing methods do not have a mechanism to promote the similarity in sparsity structures in integrative analysis. In this study, we consider penalized variable selection and estimation in integrative analysis. We develop an L₀-penalty based method, which explicitly promotes the similarity in sparsity structures. Computationally it is realized using a coordinate descent algorithm. Theoretically it has the selection and estimation consistency properties. Under a wide spectrum of simulation scenarios, it has identification and estimation performance comparable to or better than the alternatives. In the analysis of three lung cancer datasets with gene expression measurements, it identifies genes with sound biological implications and satisfactory prediction performance.