

Departmental Colloquium

Supervised Homogeneity Pursuit via Mixed Integer Optimization

Peter Song (University of Michigan)

Abstract: Stratification is one statistical principle in data processing to mitigate the underlying population heterogeneity, which is typically handled by clustering when stratum labels are unknown. Many practical problems require post-clustering statistical learning that is challenged by the issue of “double data dipping”, leading to the difficulty of uncertainty quantification. One solution to address this challenge is to perform a simultaneous operation of clustering and estimation in data analyses. Recently we developed a new paradigm of supervised homogeneity pursuit via mixed integer optimization, which provides a conceptually simple and computationally straightforward machinery with the use of suitable constraints in optimization. This novel toolbox has been then applied to solve several real-world problems arising from infectious disease surveillance, influence of environmental exposure to health, and risk factors for aging. Some algorithmic limitations worth future research will be discussed.

Friday, March 10 at 3:00 PM in 636 SEO