Statistics and Data Science Seminar

High-dimensional modeling and computation challenges and solutions via Bayesian ultrahigh dimensional variable selection and manifold-constrained optimization

Hsin-Hsiung Huang (University of Central Florida)

Abstract: High-dimensional data have become prevalent in all fields that need statistical modeling and data analysis. I introduce my recent research in Bayesian ultrahigh dimensional variable selection, low-rank matrix regression and classification, and robust sufficient dimension reduction (SDR). We develop a Bayesian framework for mixed-type multivariate regression with continuous shrinkage priors that enables joint analysis of mixed continuous and discrete outcomes, allowing variable selection from a large number of covariates (p). We investigate the conditions for posterior contraction, especially when the number of covariates (p) grows exponentially relative to the sample size (n) and develop a two-step approach for variable selection with theorems of a sure screening property and posterior contraction and applications with simulation studies and applications to real datasets.

To address challenges in analyzing regression coefficient estimation affected by high-dimensional matrix-valued covariates, we propose a framework for matrix-covariate regression and classification models with a low-rank constraint and additional regularization for structured signals, considering continuous and binary responses, introduce an efficient Riemannian-steepest-descent algorithm for regression coefficient estimation, and prove the consistency of the proposed estimator, showing

Wednesday, April 3 at 4:00 PM in 636 SEO
improvement over existing work in cases where the rank is small with applications through simulations and real datasets of shape images, brain signals, and microscopic leucorrhea images. We propose a novel SDR method robust against outliers using $\ell_2$-distance covariance that effectively estimates the central subspace under mild conditions on predictors without estimating a link function, based on the projection on the Stiefel manifold. We establish convergence properties of the proposed estimation under certain regularity conditions and compare the method's performance with existing SDR methods through simulations and real data analysis, highlighting improved computational efficiency and effectiveness.