Data Science Seminar

Hosted by Department of Mathematical Sciences

- Date: Tuesday, Oct 22, 2019
- Time: 12:00pm – 1:00pm
- Room: WH-100E
- Speaker: Gen Li (Columbia University)
- Title: Integrative multi-view regression: Bridging group-sparse and low-rank models.

Abstract

Multi-view data have been routinely collected in various fields of science and engineering. A general problem is to study the predictive association between multivariate responses and multi-view predictor sets, all of which can be of high dimensionality. It is likely that only a few views are relevant to prediction, and the predictors within each relevant view contribute to the prediction collectively rather than sparsely. We cast this new problem under the familiar multivariate regression framework and propose an integrative reduced-rank regression (iRRR), where each view has its own low-rank coefficient matrix. As such, latent features are extracted from each view in a supervised fashion. For model estimation, we develop a convex composite nuclear norm penalization approach, which admits an efficient algorithm via alternating direction method of multipliers. Extensions to non-Gaussian and incomplete data are discussed. Theoretically, we derive non-asymptotic oracle bounds of iRRR under a restricted eigenvalue condition. Our results recover oracle bounds of several special cases of iRRR including Lasso, group Lasso, and nuclear norm penalized regression. Therefore, iRRR seamlessly bridges group-sparse and low-rank methods and can achieve substantially faster convergence rate under realistic settings of multi-view learning. Simulation studies and an application in the Longitudinal Studies of Aging further showcase the efficacy of the proposed methods.

Bio: Dr. Gen Li is an assistant professor and Sanford Bolton Faculty Scholar in the Department of Biostatistics at Columbia University. His research interests include high dimensional data analysis, tensor analysis, and microbiome data analysis.