Statistics Seminar

Friday, November 12, 12:00 – 1:00 pm

Bayesian variable selection based on empirical likelihood for ultra-high dimensional data

Sherry Wang, Southern Methodist University

Abstract: A great amount of literature has shown that development of variable selection techniques can enable efficient and interpretable analysis of high dimensional data, which are ubiquitous nowadays. However, variable selection involving ultra-high dimensional data, where the number of covariates p is (much) large than the sample size n, remains a highly challenging task. Furthermore, many popular methods based on linear regression models assume Gaussian random noise. In the semi-parametric domain, under the ultra-high dimensional setting, we propose a Bayesian empirical likelihood method for variable selection, which requires no distributional assumptions but only estimating equations. Motivated by Chang et al. (2018, Annals of Statistics) on doubly penalized empirical likelihood (EL), we introduce priors to regularize both regression parameters and Lagrange multipliers associated with the estimating equations, to promote sparse learning. We further develop an efficient Markov chain Monte Carlo sampling algorithm based on the active set idea, which has been proved to be useful in reducing computational burden in several existing studies. The proposed method not only inherits merits from both Bayesian and EL inferences, but also has superior performance in both the prediction and variable selection, as shown in our numerical studies.

Bio: Dr. Xinlei (Sherry) Wang is a professor of Statistics in Department of Statistical Science, Gerald J. Ford Senior Research Fellow at Southern Methodist University. She received her B.S. degree in automatic control from University of Science and Technology of China in 1997, a Master's degree in Statistics from University of Pittsburgh in 1998, and a Ph.D. degree in Statistics from the McCombs School of Business at University of Texas at Austin in 2002. Her research interests include Bayesian hierarchical modeling, spatial modeling, statistical genomics, meta-analysis, statistics in artificial intelligence, and order-related design, theory and inference. In the past, Dr. Wang directed 16 doctoral dissertations as sole or co-mentor, as well as serving on the PhD committees of 23 other students. She has served as sole PI or co-PI in multiple grants from NIH, NSF or other agencies.