Abstract: Large samples have been generated routinely from various sources. Classic statistical models, such as smoothing spline ANOVA models, are not well equipped to analyze such large samples due to expensive computational costs. In particular, the daunting computational costs of selecting smoothing parameters render the smoothing spline ANOVA models impractical. In this talk, I will present an asypirical (asymptotic + empirical) smoothing parameters selection approach for smoothing spline ANOVA models in large samples. The proposed method can significantly reduce computational costs of selecting smoothing parameters in high-dimensional and large-scale data. We show smoothing parameters chosen by the proposed method tend to the optimal smoothing parameters minimizing a risk function. In addition, the estimator based on the proposed smoothing parameters achieves the optimal convergence rate. Extensive simulation studies will be presented to demonstrate numerical advantages of our method over competing methods. I will further illustrate the empirical performance of the proposed approach using two real data examples.

Bio: Professor Ma is a Professor of Statistics and co-directs the big data analytics lab at the University of Georgia, USA. He was Beckman Fellow at the Center for Advanced Study at the University of Illinois at Urbana-Champaign, Faculty Fellow at the US National Center for Supercomputing Applications, and a recipient of the US National Science Foundation CAREER Award. His paper won the best paper award of the Canadian Journal of Statistics in 2011. He serves on multiple editorial boards including the Journal of the American Statistical Association and Statistical Applications in Genetics and Molecular Biology. He is a Fellow of the American Statistical Association.