

Bayesian Variable Selection for Linear Regression with Interaction Terms

Wensheng Zhu

Northeast Normal University

E-mail: wszhu@nenu.edu.cn

Abstract: In modern statistics, we always encounter high-throughput data with huge number of covariates or features for a small number of subjects. In many cases, researchers show a great interest in the interactions of these covariates. However, it is a challenging task to construct the suitable model by selecting the active covariates among tens of thousands of interactions. In this talk, we propose a Bayesian variable selection approach to identify interactions in the presence of huge dimensional covariates. Spike and slab Gaussian priors are used on the main effects as well as interactions, which shrink and diffuse, respectively as the sample size increases. To reduce computational complexity, our method can be carried out by the Gibbs sampler called “Skinny Gibbs”, which was proposed recently in JASA. We show that Skinny Gibbs sampler has a stationary distribution and also exhibits the strong model selection consistency even when $p = \exp(o(n))$. Simulations in genetic association studies indicate that our proposed method offers merits in the detection of gene-gene and gene-environmental interactions.