

Statistics Seminar

Topic: Spike and slab variable selection: frequentist and Bayesian strategies
(In DNA Microarray Data Analysis)

Speaker: Dr. J. Sunil Rao
Center for Computational Genomics and System Biology,
Department of Biostatistics, Case Western Reserve
University, Cleveland, Ohio

Abstract:

Variable selection in the linear regression model takes many apparent faces from both frequentist and Bayesian standpoints. We introduce a variable selection method referred to as a rescaled spike and slab model. We study the importance of prior hierarchical specifications and draw connections to frequentist generalized ridge regression estimation. Specifically, we study the usefulness of continuous bimodal priors to model hypervariance parameters, and the effect scaling has on the posterior mean through its relationship to penalization. We demonstrate the importance of selective shrinkage for effective variable selection in terms of risk misclassification, and show this is achieved using the posterior from a rescaled spike and slab model. The method is illustrated using simulated data and a DNA microarray analysis where we search for differentially expressing genes across samples collected from different stages of colon cancer to liver metastasis. The common feature of the illustrations is the high dimensionality of the predictor space.

Date: Friday, October 8, 2004

Time: 11 AM

Location: Hume 331

ALL ARE WELCOME.

Tea/coffee/snacks will be served at 10.45 AM in the faculty lounge.