Bayesian Non-Linear Methods for Survival Analysis

and Structural Equation Models

Zhenyu Wang

Dr. Sounak Chakraborty, Co-Adviser

Dr. (Tony) Jianguo Sun, Co-Adviser

Abstract

High dimensional complex data are very common nowadays due to the technological advancement in the field of the computer science, biology, and medicine. Modeling non-linear structures in complex data and selecting the important variables or signals in a high dimensional data is a very fundamental problem in the field of statistics. In this dissertation we have developed a Bayesian kernel machine model which can incorporate existing information on pathways and gene networks in the analysis of DNA microarray data. Each pathway is modeled nonparametrically using reproducing kernel Hilbert space. The pathways and the genes are selected via assigning mixture priors on the pathway indicator variables and the gene indicator variables. This approach provides us with a flexible modeling of the pathway effects, which can capture both linear and non-linear effects. Moreover, the model can also pinpoint the important pathways and the important active genes within each pathway. As an extension to the pathway modeling, we apply the Bayesian Graph Laplacian Model which can find out the underlying gene network structure for a given set of genes. This is particularly useful when we have a set of variables or genes and we want to find out how they are connected among themselves based on a particular data. The utility and the effectiveness of our model are established by several simulation studies and real data analysis.

We extend the idea of semiparametric structural equation model where the nonlinear functional relationships are approximated using basis expansions [Guo et al., 2012]. Many basis expansion methods, including cubic splines, are known to induce correlations. In this chapter, we compare standard Lasso, Fused Lasso and Elastic Net to account for correlations in both the covariate and basis expansions. To illustrate the usefulness of the proposed methods, a simulation study and a real data study have been performed. The semiparametric structural equation models based on Bayesian fused Lasso and Bayesian elastic-net outperform the Bayesian Lasso model.