Variable Selection in Nonparametric Additive Quantile Regression for Genomic data with Priori Information

Peitao Wu, Josee Dupuis, and Ching-Ti Liu

Department of Biostatistics, Boston University School of Public Health

Abstract

A priori information, such as biological pathways, is a useful supplement in identifying risk factors of a trait using genomic data. However, the commonly used methods to incorporate prior information provide a model for the mean function of the outcome and rely on unmet assumptions. To address these concerns, we propose a method for variable selection in nonparametric additive quantile regression with network regularization to incorporate the information encoded by known networks. To account for nonlinear associations, we approximate the unknown additive functional effect of each predictor with the expansion of a B-spline basis. We implement the group Lasso penalty to obtain a sparse model. We define the network-constrained penalty by the total ℓ_2 norm of the difference between the effect functions of any two predictors that are linked in the known network. We further propose an efficient computation procedure to solve the optimization problem that arises in our model. Simulation studies show that our proposed method performs well in identifying more truly associated variables/genes and less falsely associated variables/genes than alternative approaches. We apply the proposed method to analyze the microarray gene-expression dataset in the Framingham Heart Study and identify several body mass index associated genes. In conclusion, our proposed approach efficiently identifies the outcome-associated variables in a nonparametric additive quantile regression framework by leveraging known network information.

Keywords: additive model, data integration, group Lasso, network regularization, variable selection