

SEMIPARAMETRIC REGRESSION OF MULTIDIMENSIONAL GENETIC PATHWAY DATA: LEAST-SQUARES KERNEL MACHINES AND LINEAR MIXED MODELS

Dawei Liu¹, Xihong Lin² and Debashis Ghosh³

¹*Center for Statistical Sciences, Brown University, Rhode Island, USA*

²*Department of Biostatistics, Harvard School of Public Health, Boston, USA*

³*Department of Biostatistics, University of Michigan, USA*

We consider a semiparametric regression model that relates a normal outcome to covariates and a genetic pathway, where the covariate effects are modeled parametrically and the pathway effect of multiple gene expressions is modeled parametrically or nonparametrically using least-squares kernel machines (LSKMs). This unified framework allows a flexible function for the joint effect of multiple genes within a pathway by specifying a kernel function and allows for the possibility that each gene expression effect might be nonlinear and the genes within the same pathway are likely to interact with each other in a complicated way. This semiparametric model also makes it possible to test for the overall genetic pathway effect. We show that the LSKM semiparametric regression can be formulated using a linear mixed model. Estimation and inference hence can proceed within the linear mixed model framework using standard mixed model software. Both the regression coefficients of the covariate effects and the LSKM estimator of the genetic pathway effect can be obtained using the best linear unbiased predictor in the corresponding linear mixed model formulation. The smoothing parameter and the kernel parameter can be estimated as variance components using restricted maximum likelihood. A score test is developed to test for the genetic pathway effect. Model/variable selection within the LSKM framework is discussed. The methods are illustrated using a prostate cancer data set and evaluated using simulations.