Variable selection and modeling averaging in semiparametric overdispersed generalized linear models

Remy Cottet

University of NSW

Abstract: "Flexibly modeling the response variance in regression is important for efficient parameter estimation, correct inference, and for understanding the sources of variability in the response. Our article considers flexibly modelling the mean and variance functions within the framework of double exponential regression models, a class of overdispersed generalized linear models. The most general form of our model describes the mean and dispersion parameters in terms of additive functions of the predictors. Each of the additive terms can be either null, linear, or a fully flexible smooth effect. When the dispersion model is null the mean model is linear in the predictors and we obtain a generalized linear model, whereas with a null dispersion model and fully flexible smooth terms in the mean model we obtain a generalized additive model. Whether or not to include predictors, whether or not to model their effects linearly or flexibly, and whether or not to model overdispersion at all is determined from the data using a fully Bayesian approach to inference and model selection. Model selection is accomplished using a hierarchical prior which has many computational and inferential advantages over priors used in previous empirical Bayes approaches to similar problems. We describe an efficient Markov chain Monte Carlo sampling scheme and priors that make the estimation of the model practical with a large number of predictors. The methodology is illustrated using real and simulated data."

Key Words: Double exponential regression models; Hierarchical priors; Splines; Variance estimation.