Stochastic Boosting Projected Additive Regression Spline

鄭順林
Department of Statistics, Tunghai University

Abstract

When a response has a large number of predictors (explanatory variables) and many of those predictors are redundant, model selection (also variable selections) becomes an important issue. Multivariate additive regression spline (MARS) is a powerful method to build a nonlinear model by a greedy search procedure. There are several advantages to use this method such as natural handling of data of mixed type, computational scalability for large sample size, and interpretability. However the computational effort is huge when the number of predictors is large and the boosting procedure is implemented. To overcome the pitfalls of the MARS, we propose a hybrid procedure that uses the projected predictors and the procedure of MARS.

In this research, we extend the result by Frideman (2002) and establish a procedure called "stochastic boosting projected additive regression splines"(SB-PARS). We use several data sets including hand writing numbers and microarray data to illustrate the benefits in classification ability and computational speed. We compare our proposed procedure with several sophisticated methods, such as classification and regression tree, multiple additive regression tree (MART), and stochastic MART.