

Variable Selection for Sparse High-Dimensional Nonlinear Regression Models by Combining Nonnegative Garrote and Sure Independence Screening

Shuang Wu^{1†*}, Hongqi Xue^{1†}, Yichao Wu² and Hulin Wu¹

¹Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY; ²Department of Statistics, North Carolina State University, Raleigh, NC

Version August 2013

ABSTRACT

In many regression problems, the relations between the covariates and the response may be nonlinear. Motivated by the application of reconstructing a gene regulatory network, we consider a sparse high-dimensional additive model with the additive components being some known nonlinear functions with unknown parameters. To identify the subset of important covariates, we propose a new method for simultaneous variable selection and parameter estimation by iteratively combining a large-scale variable screening (the nonlinear independence screening, NLIS) and a moderate-scale model selection (the nonnegative garrote, NNG) for the nonlinear additive regressions. We have shown that the NLIS procedure possesses the sure screening property and it is able to handle problems with non-polynomial dimensionality; and for finite dimension problems, the NNG for the nonlinear additive regressions has selection consistency for the unimportant covariates and also estimation consistency for the parameter estimates of the important covariates. The proposed method is applied to simulated data and a real data example for identifying gene regulations to illustrate its numerical performance.

Key Words: Gene regulations, independence learning, nonlinear regressions, nonnegative garrote, sigmoid function, sure screening.

*Corresponding author. †Joint first authors. *Email:* Shuang Wu, shuang.wu@urmc.rochester.edu; Hongqi Xue, hongqi_xue@urmc.rochester.edu; Yichao Wu, ywu11@ncsu.edu; Hulin Wu, hulin_wu@urmc.rochester.edu. This research was supported by the following grants: Center for Biodefense Immune Modeling, NIH/NIAID HHSN272201000055C; Center for AIDS Research, NIH/NIAID P30AI078498; Respiratory Pathogens Research Center, NIH/NIAID HHSN27220201200005C; Centers of Excellence for Influenza Research and Surveillance (CEIRS) at the University of Rochester, NIH/NIAID HHSN266200700008C; R01 AI087135; NSF grant DMS-1055210; NIH/NCI grant R01-CA149569; and two University of Rochester CTSI pilot awards (UL1RR024160) from the National Center For Research Resources.