Nonparametric Independence Screening in Sparse Ultra-High Dimensional Feature Space

Rui Song Department of Statistics Colorado State University Fort Collins CO 80526-1877 song@stat.colostate.edu

Abstract

A variable screening procedure via correlation learning was proposed in Fan and Lv (2008) to reduce dimensionality in sparse ultra-high dimensional models. Even when the true model is linear, the marginal regression can be highly nonlinear. To address this issue, we further extend the correlation learning to marginal nonparametric learning. Our nonparametric independence screening is called NIS, a specific member of the sure independence screening. Several closely related variable screening procedures are proposed. Under general nonparametric models, it is shown that under some mild technical conditions, the proposed independence screening methods enjoy a sure screening is also explicitly quantified. As a methodological extension, a data-driven thresholding and an iterative nonparametric independence screening (INIS) are also proposed to enhance the finite sample performance for fitting sparse additive models. The simulation results and a real data analysis demonstrate that the proposed procedure works well with moderate sample size and large dimension and performs better than competing methods. This is a joint work with Professor Jianqing Fan and Yang Feng.