

Model-free feature screening for ultrahigh-dimensional data conditional on some variables

Yi Liu^{1,2} · Qihua Wang^{1,3}

Received: 2 June 2016 / Revised: 18 November 2016 / Published online: 17 January 2017
© The Institute of Statistical Mathematics, Tokyo 2017

Abstract In this paper, the conditional distance correlation (CDC) is used as a measure of correlation to develop a conditional feature screening procedure given some significant variables for ultrahigh-dimensional data. The proposed procedure is model free and is called conditional distance correlation-sure independence screening (CDC-SIS for short). That is, we do not specify any model structure between the response and the predictors, which is appealing in some practical problems of ultrahigh-dimensional data analysis. The sure screening property of the CDC-SIS is proved and a simulation study was conducted to evaluate the finite sample performances. Real data analysis is used to illustrate the proposed method. The results indicate that CDC-SIS performs well.

Keywords Conditional distance correlation · Feature selection · Sure screening property · High-dimensional data

Electronic supplementary material The online version of this article (doi:[10.1007/s10463-016-0597-2](https://doi.org/10.1007/s10463-016-0597-2)) contains supplementary material, which is available to authorized users.

✉ Qihua Wang
qhwang@amss.ac.cn

¹ Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

² College of Science, China University of Petroleum, Qingdao 266580, China

³ Institute of Statistical Science, Shenzhen University, Shenzhen 518006, China