

TWO KERNEL-BASED FEATURE SCREENING PROCEDURES FOR HIGH-DIMENSIONAL RESPONSE DATA

Yuke Shi^{1,2}, Na Li³, Qizhai Li^{1,2}, Dongdong Pan⁴ and Jinjuan Wang^{*5}

¹*Chinese Academy of Sciences*, ²*University of Chinese Academy of Sciences*,
³*Beijing Information Science and Technology University*,
⁴*Yunnan University* and ⁵*Beijing Institute of Technology*

Abstract: We consider feature screening for high-dimensional response data without and with the existence of confounding factors. First, we introduce kernel covariance and kernel correlation for high-dimensional association analysis, and further propose partial kernel covariance and partial kernel correlation that can handle situations with confounding factors. Then, based on the kernel correlation and partial kernel correlation, we propose two feature screening procedures. Both screening procedures possess sure screening property and ranking consistency property, and are complementary to each other by respectively dealing with situations without and with the existence of confounding factors. The proposed procedures make no assumptions on model, and are suitable for high-dimensional response variable and non-Euclidean data. Extensive simulation results and a real data analysis demonstrate the satisfying performances and advantages of the proposed procedures over existing methods.

Key words and phrases: Confounding factors, feature screening, high-dimensional response variable, kernel correlation, partial kernel correlation.

1. Introduction

High-dimensional responses are frequently encountered in feature screening analysis nowadays. For example, in both gene expression experiments and genetic pleiotropic association studies that often search for disease-associated genetic variants among hundreds of thousands of variants, the responses are usually high-dimensional, which are expression levels of tens of thousands of genes in the former analysis and are hundreds of human complex traits in the latter study, respectively (Gratten and Visscher, 2016; Zhang et al., 2017; Watanabe et al., 2019).

A good strategy to complete such searching is feature screening accompanied with association analysis. There exist many variable screening procedures dealing with diverse situations. For example, feature screening procedures for linear model (Fan and Lv, 2008), generalized linear model (Fan and Song, 2010),

*Corresponding author. E-mail: wangjinjuan@bit.edu.cn