



Surrogate-variable-based model-free feature screening for survival data under the general censoring mechanism

Jing Zhang¹ · Qihua Wang^{2,3} · Xuan Wang⁴

Received: 20 May 2020 / Revised: 26 March 2021 / Accepted: 12 April 2021 /
Published online: 3 June 2021
© The Institute of Statistical Mathematics, Tokyo 2021

Abstract

Feature screening has been seen as the first step in analyzing the ultrahigh-dimensional data with the censored survival time. In this article, we develop a surrogate-variable-based model-free feature screening approach for the censored data under the general censoring mechanism, where the censoring variable may depend on the survival variable and the covariates. This approach is developed by finding some observable variables whose active covariates contain the active covariates of the survival variable as a subset, respectively. Then, any existing model-free feature screening method with the sure screening property for full data can be applied to estimating the sets of the active covariates of the observable variables and hence the set of the active covariates of the survival variable. The sure screening property of the proposed approach is established, and its finite sample performances are demonstrated through some simulations. Further, we illustrate the proposed approach by analyzing two real datasets.

Keywords Feature screening · Model-free · Sure screening property · Survival data · Ultrahigh dimensionality

✉ Qihua Wang
qhwang@amss.ac.cn

Jing Zhang
jingzhang@amss.ac.cn

Xuan Wang
wangxuan209@hotmail.com

- ¹ School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China
- ² Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
- ³ School of Statistics and Mathematics, Zhejiang Gongshang University, Zhejiang 310018, China
- ⁴ School of Mathematical Sciences, Zhejiang University, Zhejiang 310018, China