



Robust high-dimensional regression for data with anomalous responses

Mingyang Ren^{1,2} · Sanguo Zhang^{1,2} · Qingzhao Zhang³

Received: 10 March 2020 / Revised: 28 June 2020 / Published online: 30 September 2020
© The Institute of Statistical Mathematics, Tokyo 2020

Abstract

The accuracy of response variables is crucially important to train regression models. In some situations, including the high-dimensional case, response observations tend to be inaccurate, which would lead to biased estimators by directly fitting a conventional model. For analyzing data with anomalous responses in the high-dimensional case, in this work, we adopt γ -divergence to conduct variable selection and estimation methods. The proposed method possesses good robustness to anomalous responses, and the proportion of abnormal data does not need to be modeled. It is implemented by an efficient coordinate descent algorithm. In the setting where the dimensionality p can grow exponentially fast with the sample size n , we rigorously establish variable selection consistency and estimation bounds. Numerical simulations and an application on real data are presented to demonstrate the performance of the proposed method.

Keywords Anomalous responses · Robust · γ -divergence · High-dimensional data

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10463-020-00764-1>) contains supplementary material, which is available to authorized users.

✉ Qingzhao Zhang
qzzhang@xmu.edu.cn

- ¹ School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
- ² Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100049, China
- ³ Department of Statistics, School of Economics, The Wang Yanan Institute for Studies in Economics, MOE Key Lab of Economics and Fujian Key Lab of Statistics, Xiamen University, Xiamen 361005, China