Ren, Mingyang; Zhang, Sanguo; Zhang, Qingzhao

Robust high-dimensional regression for data with anomalous responses. (English)


Summary: 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 $\gamma$-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.

MSC:

62R07 Statistical aspects of big data and data science
62G08 Nonparametric regression and quantile regression
62G35 Nonparametric robustness

Keywords:
anomalous responses; robustness; $\gamma$-divergence; high-dimensional data

Full Text: DOI

References:


[23] Li, Z., Expression and clinical significance of androgen receptor in triple negative breast cancer, Cancers, 9, 1, 585-590 (2017)


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original
paper as accurately as possible without claiming the completeness or perfect precision of the matching.