Oda, Ryoya

Consistent variable selection criteria in multivariate linear regression even when dimension exceeds sample size. (English) Zbl 1469.62423


Summary: This paper is concerned with the selection of explanatory variables in multivariate linear regression. The Akaike’s information criterion and the $C_p$ criterion cannot perform in high-dimensional situations such that the dimension of a vector stacked with response variables exceeds the sample size. To overcome this, we consider two variable selection criteria based on an $L_2$ squared distance with a weighted matrix, namely the scalar-type generalized $C_p$ criterion and the ridge-type generalized $C_p$ criterion. We clarify conditions for their consistency under a hybrid-ultra-highdimensional asymptotic framework such that the sample size always goes to infinity but the number of response variables may not go to infinity. Numerical experiments show that the probabilities of selecting the true subset by criteria satisfying consistency conditions are high even when the dimension is larger than the sample size. Finally, we illuminate the practical utility of these criteria using empirical data.

MSC:

62R07 Statistical aspects of big data and data science
62J05 Linear regression; mixed models
62H12 Estimation in multivariate analysis

Keywords:

hybrid-ultra-high-dimensional asymptotic framework; multivariate linear regression; non-normality; selection consistency; variable selection criterion

Full Text: DOI

References:

and its applications. J. Multivariate Anal., 130, 27-44. · Zbl 1292.62043


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.