

Performances comparison of information criterion for outlier detection in multiple regression models having multicollinearity problems using genetic algorithms

Özlem Gürünlü Alma

Mugla University, Turkey

Abstract

Multiple linear regression models are widely used applied statistical techniques and they are most useful devices for extracting and understanding the essential features of datasets. However, in multiple linear regression models problems arise when a serious outlier observation or multicollinearity present in the data. Multicollinearity is a linear dependency between two or more explanatory variables in the regression models which can seriously disturb the least squares estimated regression surface. The other important problem is outlier; they can strongly influence the estimated model, especially when using least squares method. Nevertheless, outlier data are often the special points of interests in many practical situations. The purpose of this study is to performances comparison of Akaike Information Criterion, Bayesian Information Criterion and Information Complexity criterion for detecting outliers using Genetic Algorithms when multiple regression models having multicollinearity problems.

Keywords

Akaike information criterion, Bayesian information criterion and information complexity criterion, Genetic algorithms, Multicollinearity, Multiple linear regression, Outlier detection.

References

- Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, USA.
- Gürünlü Alma, Ö., Kurt, S., and Ugur, A. (2009). Genetic algorithms for outlier detection in multiple regression with different information criteria. *J. Stat. Comput. Simul.*, Doi: 10.1080/00949650903136782.

Hoeting, J., Raftery, A.E., and Madigan, D. (1996). A method for simultaneous variable selection and outlier identification in linear regression. *Comput. Statist. Data Anal.* 22, 251–270.

Papadimitriou, S., Kitawaga, H., Gibbons, P.G., and Faloutsos, C. (2002). LOCI: Fast outlier detection using the local correlation integral. *Intel research Laboratory Technical report no. IRP-TR-02-09*.

Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *Soft Computing* 8, 527–533.

Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. In Multivariate Analysis. In Krishnaiah, P.R. (Eds.) *Multivariate Analysis* (pp.391–420). Academic Pres, New York.