Two steps robust Mahanalobis method for identification of multiple high leverage points

ABSTRACT

Problem statement: High leverage points are extreme outliers in the X-direction. In regression analysis, the detection of these leverage points becomes important due to their arbitrary large effects on the estimations as well as multicollinearity problems. Mahalanobis Distance (MD) has been used as a diagnostic tool for identification of outliers in multivariate analysis where it finds the distance between normal and abnormal groups of the data. Since the computation of MD relies on nonrobust classical estimations, the classical MD can hardly detect outliers accurately. As an alternative, Robust MD (RMD) methods such as Minimum Covariance Determinant (MCD) and Minimum Volume Ellipsoid (MVE) estimators had been used to identify the existence of high leverage points in the data set. However, these methods tended to swamp some low leverage points even though they can identify high leverage points correctly. Since, the detection of leverage points is one of the most important issues in regression analysis, it is imperative to introduce a novel detection method of high leverage points. Approach: In this study, we proposed a relatively new two-step method for detection of high leverage points by utilizing the RMD (MVE) and RMD (MCD) in the first step to identify the suspected outlier points. Then, in the second step the MD was used based on the mean and covariance of the clean data set. We called this method two-step Robust Diagnostic Mahalanobis Distance (RDMDTS) which could identify high leverage points correctly and also swamps less low leverage points. Results: The merit of the newly proposed method was investigated extensively by real data sets and Monte Carlo Simulations study. The results of this study indicated that, for small sample sizes, the best detection method is (RDMDTS) (MVE)-mad while there was not much difference between (RDMDTS) (MVE)-mad and (RDMDTS) (MCD)-mad for large sample sizes. Conclusion/Recommendations: In order to swamp less low leverage as high leverage point, the proposed robust diagnostic methods, (RDMDTS) (MVE)-mad and (RDMDTS) (MCD)-mad were recommended.

Keyword: Multicollinearity; High leverage points; Robust mahalanobis distance; Distance-distance; Plot