

The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression

ABSTRACT

Leverage values are being used in regression diagnostics as measures of influential observations in the X -space. Detection of high leverage values is crucial because of their responsibility for misleading conclusion about the fitting of a regression model, causing multicollinearity problems, masking and/or swamping of outliers, etc. Much work has been done on the identification of single high leverage points and it is generally believed that the problem of detection of a single high leverage point has been largely resolved. But there is no general agreement among the statisticians about the detection of multiple high leverage points. When a group of high leverage points is present in a data set, mainly because of the masking and/or swamping effects the commonly used diagnostic methods fail to identify them correctly. On the other hand, the robust alternative methods can identify the high leverage points correctly but they have a tendency to identify too many low leverage points to be points of high leverages which is not also desired. An attempt has been made to make a compromise between these two approaches. We propose an adaptive method where the suspected high leverage points are identified by robust methods and then the low leverage points (if any) are put back into the estimation data set after diagnostic checking. The usefulness of our newly proposed method for the detection of multiple high leverage points is studied by some well-known data sets and Monte Carlo simulations.

Keyword: Diagnostic-robust generalized potentials; Group deletion; High leverage points; Masking; Robust Mahalanobis distance; Minimum volume ellipsoid; Monte Carlo simulation