Selective overview of forward selection in terms of robust correlations

ABSTRACT

Forward selection (FS) is a very effective variable selection procedure for selecting a parsimonious subset of covariates from a large number of candidate covariates. Detecting the type of outlying observations, such as vertical outliers or leverage points, and the FS procedure are inseparable problems. For robust variable selection, a crucial issue is whether the outliers are univariate, bivariate, or multivariate. This paper uses a consistent robust multivariate dispersion estimator to obtain robust correlation estimators used to establish robust forward selection (RFS) procedures that outperform methods that use robust bivariate correlations. The usefulness of our proposed procedure is studied with a numerical example and a simulation study. The result shows the proposed method has scalability and the ability to deal with univariate, bivariate and multivariate outlying observations including leverage points or vertical outliers, and the new method outperforms previously published methods of RFS.