Abstract

Outlier detection is a fundamental step in the model construction process. This problem has attracted many researchers for a long time. Consequently, numerous methods are available in the literature. On way to tackle the problem is through feature selection approach. The problem of outlier detection in the linear regression can be easily transferred to a feature selection problem through some matrix operations. Applying feature selection method, the relevant features selected are considered as outliers. In the introduction, some basic concepts about outlier detection and feature selection are provided. Specifically, the penalized regression methods are discussed in details. Since many approaches to outlier detection are available, a literature review is conducted on some frequently cited methods and their merits and shortcomings are summarized. In this article, we take advantage of Sequential Lasso cum extended Bayesian criterion (SLasso cum EBIC) as the feature selection method in the sparse high-dimensional model. This approach is said to perform well in the high-dimensional feature space (Luo and Chen 2014). Comprehensive simulation studies comparing SLasso cum EBIC and another feature selection method called Θ-IPOD is provided to reveal their relative advantages and disadvantages. Positive discovery rate and False discovery rate are taken as measurements for their performance. The result shows that Θ-IPOD behaves quite stable among all the simulation settings, while Slasso cum EBIC has advantages under some simulation settings. Thus, the ideal choice of method depends on the purpose and nature of study.