

Detection of outliers in high-dimensional data using nu-support vector regression

Support Vector Regression (SVR) is gaining in popularity in the detection of outliers and classification problems in high-dimensional data (HDD) as this technique does not require the data to be of full rank. In real application, most of the data are of high dimensional. Classification of high-dimensional data is needed in applied sciences, in particular, as it is important to discriminate cancerous cells from non-cancerous cells. It is also imperative that outliers are identified before constructing a model on the relationship between the dependent and independent variables to avoid misleading interpretations about the fitting of a model. The standard SVR and the μ - ϵ -SVR are able to detect outliers; however, they are computationally expensive. The fixed parameters support vector regression (FP- ϵ -SVR) was put forward to remedy this issue. However, the FP- ϵ -SVR using ϵ -SVR is not very successful in identifying outliers. In this article, we propose an alternative method to detect outliers i.e. by employing nu-SVR. The merit of our proposed method is confirmed by three real examples and the Monte Carlo simulation. The results show that our proposed nu-SVR method is very successful in identifying outliers under a variety of situations, and with less computational running time.

Keyword: High-dimensional data; Outliers; Robustness; Statistical learning theory; Support vector regression