Outlier Identification in Joint Dimension Reduction and Clustering

Angelos Markos[1]; Alfonso Iodice D'Enza[2]; Michel van de Velden[3]

[1]   Democritus University of Thrace, Alexandroupoli, Greece
[2]   Federico II University of Naples, Naples, Italy
[3]   Erasmus University Rotterdam, Netherlands

Abstract:
Joint dimension reduction and clustering (JDR) refers to a class of non-model based methods that perform simultaneous dimension reduction and clustering of continuous, categorical or mixed-type data. The main advantage of JDR is that both clustering of objects and a low-dimensional subspace reflecting the cluster structure are simultaneously obtained. Although these methods are especially suited for high-dimensional data, even a small number of multivariate outliers can influence the cluster structure and the number of clusters. Outliers are here defined as data points that do not belong to any of the clusters or lie between the clusters, particularly if they are well-separated. This work generalizes a recently proposed outlier detection approach in the low dimensional space, specifically intended to identify observations that deviate from the assigned centroids, so that they can be considered external to the obtained clusters. Results from a simulation study demonstrate the ability of the method to detect outlying observations and its comparable performance to other approaches, under most scenarios.