

A computational method with density based clustering approach for the data in the existence of outlier

Fatma Yerlikaya-Özkurt

Atılım University, Ankara, Turkey

Abstract

Recently, collection of huge amount of data and analysis of that much data have vital importance for human activities in many different application areas. Advanced statistical methods play crucial role for modeling of such data when the data contains outliers. Although there are number of outlier detection methods for revealing outlier observations in data, most of them may not be reasonable and appropriate for prediction purposes due to structural and requirements of modeling. In this study, density based clustering algorithms are considered in order to detect the location of outlier observations effectively with respect to form of the model in high-dimensional data. Based on obtained results, the Mean Shift Outlier Model (MSOM) is constructed as a robust linear model. This newly proposed computational method use power of data clustering and also minimize the impact of the outlier observations. The numerical example is also presented to reveal the performance of the method in this study.

Keywords

Linear models, Outlier observation, High-dimensional data, Density based clustering.

References

- [1] Gan, G., Ma, C. and Wu, J. (2020). *Data clustering: theory, algorithms, and applications*. SIAM.
- [2] Xia, J., Gao, L., Kong, K., Zhao, Y., Chen, Y., Kui, X., and Liang, Y. (2018). Exploring linear projections for revealing clusters, outliers, and trends in subsets of multi-dimensional datasets. *Journal of Visual Languages and Computing*, 48 52–60.
- [3] Xu, R. and Wunsch, D. (2008). *Clustering*. John Wiley & Sons.
- [4] Xu, X., Liu, H., Li, L., and Yao, M. (2018). A comparison of outlier detection techniques for high-dimensional data. *International Journal of Computational Intelligence Systems*, 11(1), 652–662.