DETERMINING THE NUMBER OF CLUSTERS BY MULTI DIMENSIONAL SCALING

Murat Erişoğlu[1] and Serkan Akoğul[2]

[1] Necmettin Erbakan University, Turkey
[2] Yldz Technical University, Turkey

Clustering is a popular tool for exploratory data analysis. In cluster analysis, identifying the number of clusters in a dataset is one of the most important problems. In this study, a new approach is proposed for determining the number of clusters based on multidimensional scaling. The proposed method is easy and is computationally efficient and straightforward to estimate the number of clusters.

Keywords: Number of clusters, dimension reduction, multi-dimensional scaling.

References:

Asuncion, A. and Newman, D.J., UCI Machine Learning Repository Irvine, CA: University of California, School of Information and Computer Science, 2007.

Calinski, T. and Harabasz, J., A Dendrite Method for Cluster Analysis, Communications in Statistics, 3, 1-27, 1974.

Catherine, A.S. and Gareth, M.J., Finding the Number of Clusters in a Dataset: An Information-Theoretic Approach, J. Amer. Statist. Assoc., 98(463), 750-763, 2003.

Haojun S., Shengrui W. and Qingshan J., A New Validation Index for Determining the Number of Clusters in A Data Set, Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference Vol. 3, 852-1857, 2001.

Erişoğlu, M., Erişoğlu, Ü., Servi, T., Sakalloğlu, S., A New Approach For Determining Number Of Clusters, Pakistan Journal of Statistics, 28 (1), 141-158, 2012.