# A new algorithm for initial cluster centers in K-means clustering

## Murat Erişoğlu, Nazif Çalış, and Sadullah Sakalloğlu

Çukurova University, Turkey

#### Abstract

Clustering is a very well known technique in data mining. One of the most widely used clustering techniques is the k-means clustering. Solutions obtained from k-means clustering are dependent on the initialization of cluster centers. In this paper, we propose an algorithm to compute initial cluster centers for K-means clustering. Firstly two principal variables are selected by the proposed methods. The data set is partitioned one at a time until the number of cluster equals to the predefined number of clusters. The algorithm has been applied to different data sets and good results are obtained.

### **Keywords**

K-means algorithm, Initial cluster centers, Clustering algorithm, Rand index, Classification accuracy.

# References

Blake, C.L. and Merz, C.J. (1998). UCI Repository of machine learning databases. University of California, Irvine, Department of Information and Computer Science.

Bradley, P.S. and Fayyad, U.M. (1998). Refinning initial points for kmeans clustering. In: *Proceeding of The 15. International Conference on Machine Learning.* 

Deelers, S. and Auwatanamongkol, S. (2007). Enhancing k-means algorithm with initial cluster centers derived from data partitioning along the data axis with the highest variance. *Int. J. Computer Science* 2(4), 247–252.

Jain, A.K. and Dubes, R.C. (1988). *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ.

Khan, S.S. and Ahmad, A. (2004). Cluster center initilization algorithm for k-means clustering. *Pattern Recognition Letters* 25, 1293–1302.

Likas, A., Vlassis, N., and Jakob, J.V. (2003). The global k-means clustering algorithm. *Pattern Recognition* 36, 451–461.

Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In Cam, L.M.L., Neyman, J. (Eds.) *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California. Pres., 281–297.

Rand, W.M. (1971). Objective criteria for the evaluation of clustering methods, J. Amer. Statist. Assoc. 66, 846–850.

Ruspine, E.H. (1970). Numerical methods for fuzzy clustering. *Inform.* Sci. 2, 319–350.

2