Scaldelai, D.; Matioli, L. C.; Santos, S. R.; Kleina, M.

MulticlusterKDE: a new algorithm for clustering based on multivariate kernel density estimation. (English) Zbl 07484712

J. Appl. Stat. 49, No. 1, 98-121 (2022)

Summary: In this paper, we propose the MulticlusterKDE algorithm applied to classify elements of a database into categories based on their similarity. MulticlusterKDE is centered on the multiple optimization of the kernel density estimator function with multivariate Gaussian kernel. One of the main features of the proposed algorithm is that the number of clusters is an optional input parameter. Furthermore, it is very simple, easy to implement, well defined and stops at a finite number of steps and it always converges regardless of the data set. We illustrate our findings by implementing the algorithm in R software. The results indicate that the MulticlusterKDE algorithm is competitive when compared to K-means, K-medoids, CLARA, DBSCAN and PdfCluster algorithms. Features such as simplicity and efficiency make the proposed algorithm an attractive and promising research field that can be used as basis for its improvement and also for the development of new density-based clustering algorithms.

MSC:

62Pxx Applications of statistics

Keywords:

kernel density estimation; Gaussian kernel; clustering data; optimization method; multiclusterKDE

Software:

mclust; ClusterKDE; clusfind; Silhouettes; UCI-ml; MulticlusterKDE; pdfCluster; R

Full Text: DOI

References:


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.