

BISECTING COLLINEAR CLUSTERING ALGORITHM

TERENCE JOHNSON

Department of Computer Engineering, Agnel Institute of Technology and Design, Goa, India

ABSTRACT

The motivation for the Bisecting Collinear Clustering Algorithm comes from the Bisecting K-means Algorithm. To obtain T clusters, split the set of all points into two clusters, select one of these clusters to split (Each time Bisect the selected cluster using the Collinear Clustering Algorithm), and so on, until T clusters have been produced. Clustering, using the classical K-Means method results in obtaining final fixed points which we call the final unchanging means around which all other points in the dataset get clustered. This suggests that if we identify points in any dataset which represent the final unchanging means, then the task of clustering reduces to just assigning the remaining points in the dataset into clusters, which are closest to these final means based on standard distance measures. Taking a cue from the result of the K-Means method, the collinear clustering algorithm groups collinear data points of any dataset into different clusters, depending on which points in the dataset lie at maximum distance from each other. The clusters are formed by finding the minimum Euclidean distance of all points in the dataset and these maximally separated data points.

KEYWORDS: Bisecting Collinear Clustering, Maximal Distance Clustering, Minimum Euclidean Distance