Hafiz ahmad

19l-1316

EE#8A

Assignment # 3:

K-implies is a centroid-based grouping calculation that segments a dataset into k bunches, where every information point has a place with the group whose centroid is nearest to it. After assigning each data point to the centroid closest to it at random, the algorithm recalculates the centroids using the mean of the data points in each cluster. Until convergence, this process is repeated.

DBSCAN, on the other hand, is a density-based clustering algorithm that separates data points in low-density areas from those in high-density areas that are close to one another. Two parameters govern how the algorithm operates: epsilon, which is the sweep of the area around a relevant piece of information, and minPts, which is the base number of focuses expected to shape a thick locale. DBSCAN begins by choosing an irregular data of interest and finding every one of the focuses inside epsilon distance. On the off chance that the quantity of focuses is more prominent than or equivalent to minPts, another group is shaped. The algorithm then adds all points that are also within epsilon distance and have not yet been assigned to a cluster, expanding the cluster.

Silhouette score:

The silhouette score can be used to compare how well the DBSCAN and k-means algorithms perform and decide which one is better for a given task. A metric for assessing the quality of clustering results is the silhouette score. It shows how distinct clusters are from one another and how well data points fit into the clusters that have been assigned to them. A summary of the silhouette score is as follows:

The silhouette score is between -1 and 1, with:

• A score close to 1 indicates that the data points are tightly grouped within their respective clusters and well-separated clusters.

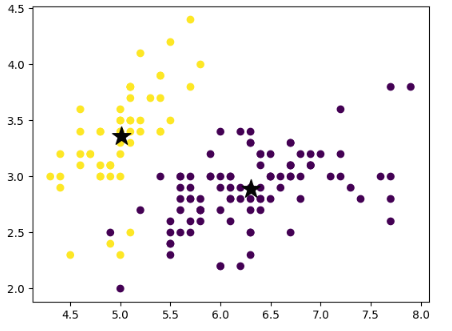
• A score around 0 recommends covering or equivocal groups, where information focuses may not be plainly doled out to a particular bunch.

• A score close to -1 indicates that the clustering structure is incorrect or that the data points were likely assigned to the wrong clusters.

The difference between the average distance to data points in its own cluster (cohesion) and the average distance to data points in the closest neighboring cluster (separation) is used to calculate the silhouette score for each data point. The general outline score is the normal of the outline scores of all data of interest in the dataset.

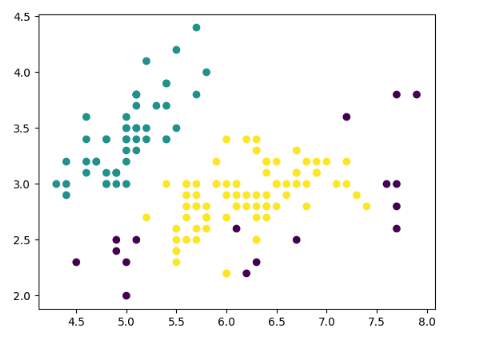
A higher silhouette score indicates better clustering performance when comparing different clustering algorithms or evaluating various parameter settings. A summary of how to use the silhouette score to compare DBSCAN and k-means is as follows:

* Apply the dataset to both the DBSCAN and k-means algorithms.
* Compute the outline score for each grouping arrangement acquired from the two calculations.
* Compare the scores of the silhouette produced by the two algorithms. A higher outline score shows better bunching execution.
* Choose the algorithm with the highest silhouette score as the best option for your particular task based on the comparison.
* K-Means Algorithm on the IRIS dataset:





* DBSCAN Algorithm on the IRIS dataset:





The silhouette score for the k-means algorithm on the iris dataset is 0.6808, while the silhouette score for the DBSCAN algorithm on the same dataset is 0.4858, as shown in the attached snapshots of the results. The k-means algorithm outperforms the DBSCAN algorithm on the iris dataset, which confirms my earlier statement that a higher silhouette score indicates superior performance when comparing two clustering algorithms.