

DB Scan

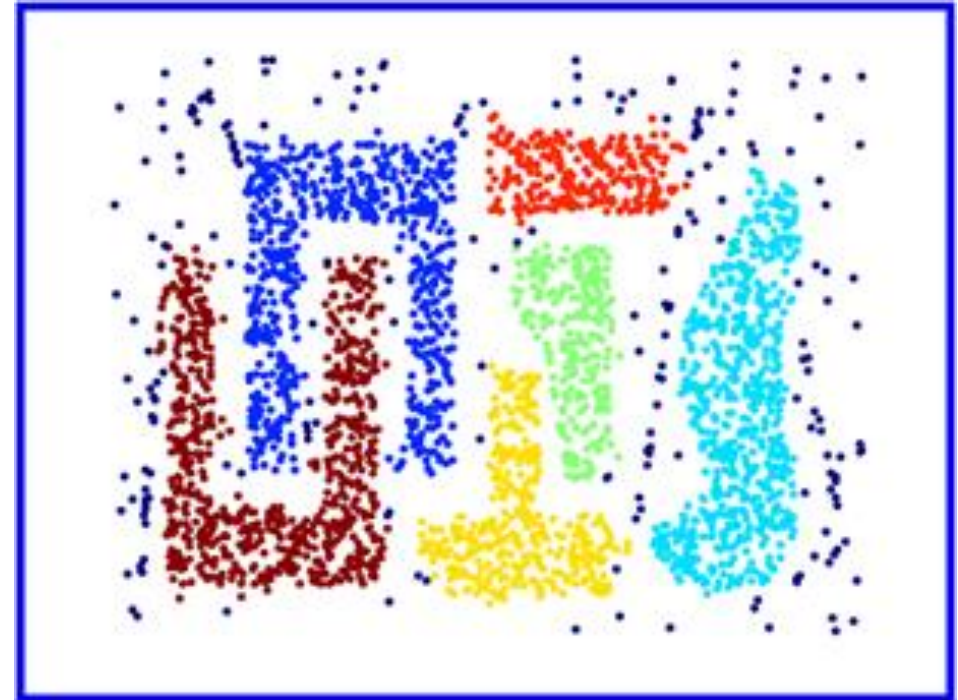
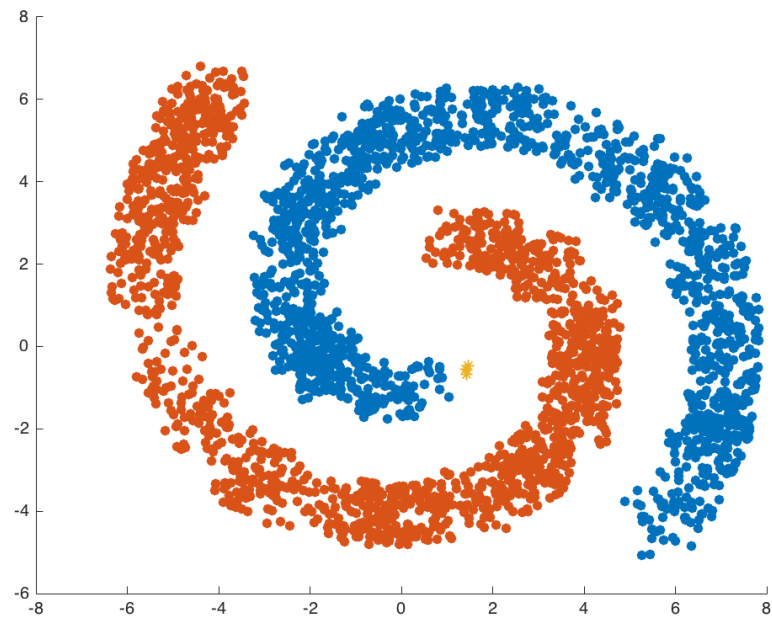
BY MG ANALYTICS

DB Scan

- ▶ Takes 2 Parameters:
- ▶ Epsilon – neighborhood size
- ▶ Min points
- ▶ Autodetects number of cluster.
- ▶ Does not make assumption of spherical clusters
- ▶ Can be used for anomaly detection.
- ▶ Does not get impacted by outliers

Link to try.

- ▶ <https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/>



DBSCAN



k-means



Calculation of Silhouette Value –

If the Silhouette index value is high, the object is well-matched to its own cluster and poorly matched to neighbouring clusters. The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient is defined as –

$$S(i) = (b(i) - a(i)) / (\max \{ (a(i), b(i)) \})$$

Where,

- a(i) is the average dissimilarity of i^{th} object to all other objects in the same cluster
- b(i) is the average dissimilarity of i^{th} object with all objects in the closest cluster.

Range of Silhouette Value –

Now, obviously $S(i)$ will lie between $[-1, 1]$ –

1. If silhouette value is close to 1, sample is well-clustered and already assigned to a very appropriate cluster.
2. If silhouette value is about 0, sample could be assigned to another cluster closest to it and the sample lies equally far away from both the clusters. That means it indicates overlapping clusters
3. If silhouette value is close to -1 , sample is misclassified and is merely placed somewhere in between the clusters.