

CLUSTERING ALGORITHMS IN MACHINE LEARNING

SWIPE →



01/K-Means Clustering

It can group an unlabeled dataset very quickly.

It works by labelling all instances on the cluster with the closest centroid.

It starts with placing the centroids randomly by selecting k random instances and using their locations as the centroids. Then you can label the instances, update the centroids, and so on.

The K-Means clustering algorithm is guaranteed to converge in a few iterations. It will not continue to iterate forever.

02/DBSCAN Clustering

It's based on these concepts:

- **Core Samples:** The samples present in the high-density area have minimum sample points with the eps radius.
- **Non-core samples:** These lie within the eps radius of the core samples but they don't have minimum samples points.
- **Outliers:** The samples that are not part of the core samples and the non-core samples and are far away from all the samples.

The DBSCAN clustering algorithm works well if all the clusters are dense enough and are well represented by the low-density regions.

03/Agglomerative Clustering

Here the process of grouping similar instances starts by creating multiple groups where each group contains one entity at the initial stage, then it finds the two most similar groups, merges them, repeats the process until it obtains a single group of the most similar instances.

Some of the advantages of using this algorithm for clustering are:

- It adapts very well to a large number of instances
- It can capture the clusters of different shapes
- It forms flexible and informative clusters
- It can also be used with any pairwise distance

04/BIRCH Clustering

It's specially designed for clustering on a very large dataset.

It is often faster than other clustering algorithms. It provides a very similar result to the K-Means algorithm if the number of features in the dataset is not more than 20.

When training the model using the BIRCH algorithm, it creates a tree structure with enough data to quickly assign each data point to a cluster.

By storing all the data points in the tree, this algorithm allows the use of limited memory while working on a very large dataset.

05/Mean Shift Clustering

The complete process of the Mean Shift clustering algorithm:

- It starts by placing a circle centered on each sample
- Then for each circle, it calculates the mean of all the samples located in the circle
- Then it moves the circle so that it is centered on the mean
- Then it iterates the mean shift step until all of the circles stop moving
- Then it shifts the circles in the direction of the highest density until each circle reaches a maximum of local density
- Then all the instances whose circles have settled in the same place are assigned to the same cluster



FOUND IT HELPFUL?

Let us know in the comments

At DDS, we will help you build a career in Data Science

[LINK IN BIO](#)
