**CLUSTERING**

Without going into the technicalities of the definition of clustering first, what I basically understood based on reading the PDFs was that one of the primary motivations behind clustering is to analyse the data by segragating in the form of clusters or groups. For example if a company wants to analyse what kind of customers like what kind of the product, It will divide its entire customer into two or more clusters, where each cluster or group signifies or represents that group liking a particular kind of product. Hence, once the understanding of company’s customer base is clear, products can be developed according to the preferences of each and in a better manner.

So basically we can conclude that Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to each other in the same group than those in other groups. In simple words, the aim is to divide groups with similar behaviour and assign them into clusters. The process of clustering is of great usefulness because through the process of clustering the data, we can obtain the data distribution, observe the characteristics of each cluster and make further studies on any particular cluster or all the clusters.

## 🡪 A well-separated cluster in which a cluster is a set of points such that any point that is in a cluster is closer (or more similar) to every other point in the cluster than to any point which is not in the cluster.

## Another one is a

## 🡪Centre-based cluster a cluster is a set of objects such that an object in a cluster is closer (more similar) to the “centre” of a cluster, than to the centre of any other cluster. The centre of a cluster is often a the average of all the points in the cluster or the “most representative” point of a cluster.

## K Means Clustering

There are a number of clustering algorithms but one of the most prominent one id K-Means clustering.

It is basically an iterative algorithm that that continues until no further changes are possible in the clusters formed. A basic outline of the steps followed in K means clustering are:

--To begin with, take K as the input which is how many clusters you want to find. Place K points, which serves as the centroids, in random locations in the space.

--Next step involves using the distance between data points and centroids, assign each data point to the cluster which is close to it.

--Calculate again, the cluster centers as a mean of data points assigned to it.

--We shall then repeat steps 2 and 3 until no further changes are observed in the way the clusters are formed.

Regards

Ayush Anand

3rd Year Undergraduate

Information Technology

IIEST Shibpur