

Machine Learning Solution

Objective Questions

Q1.

Ans: b) 4

Q2.

Ans: d) 1, 2 and 4

Q3.

Ans: a) interpreting and profiling clusters

Q4.

Ans: a) Euclidean distance

Q5.

Ans: b) Divisive clustering

Q6.

Ans: d) All answers are correct

Q7.

Ans: a) Divide the data points into groups

Q8.

Ans: a) Supervised learning

Q9.

Ans: c) Diverse clustering

Q10.

Ans: a) K-means clustering algorithm

Q11.

Ans: d) All of the above

Q12.

Ans: b) Unlabelled data

Subjective Questions

Q13.

Ans:

Q14.

Ans: A cluster quality is measured through its performance, the capability to find out hidden patterns, and the similarity measure utilized by the methodology.

Q15.

Ans: Cluster Analysis:

Cluster analysis is a multivariate data mining technique whose goal is to group objects (eg., products, respondents, or other entities) based on a set of user selected characteristics or attributes. It is the basic and most important step of data mining and a common technique for statistical data analysis, and it is used in many fields such as data compression, machine learning, pattern recognition, information retrieval etc.

Types of Cluster analysis:

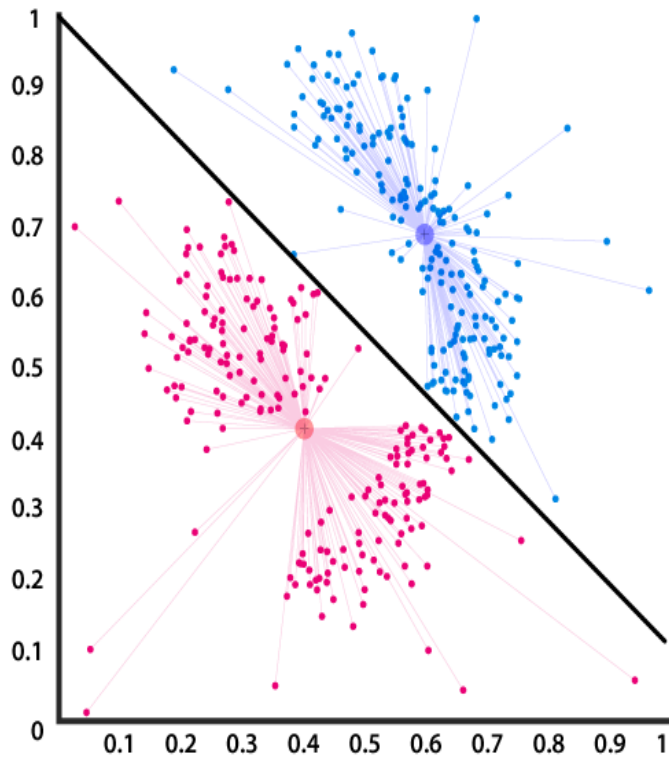
Hierarchical Cluster Analysis

In this method, first, a cluster is made and then added to another cluster (the most similar and closest one) to form one single cluster. This process is repeated until all subjects are in one cluster. This particular method is known as Agglomerative method. Agglomerative clustering starts with single objects and starts grouping them into clusters.

The divisive method is another kind of Hierarchical method in which clustering starts with the complete data set and then starts dividing into partitions.

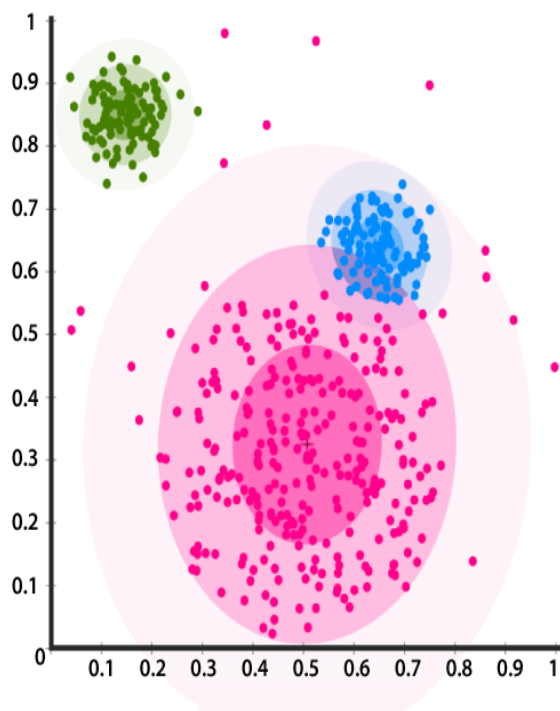
Centroid-based Clustering

In this type of clustering, clusters are represented by a central entity, which may or may not be a part of the given data set. K-Means method of clustering is used in this method, where k are the cluster centres and objects are assigned to the nearest cluster centres.



Distribution-based Clustering

It is a type of clustering model closely related to statistics based on the modals of distribution. Objects that belong to the same distribution are put into a single cluster. This type of clustering can capture some complex properties of objects like correlation and dependence between attributes.



Density-based Clustering

In this type of clustering, clusters are defined by the areas of density that are higher than the remaining of the data set. Objects in sparse areas are usually required to separate clusters. The objects in these sparse points are usually noise and border points in the graph. The most popular method in this type of clustering is DBSCAN.

