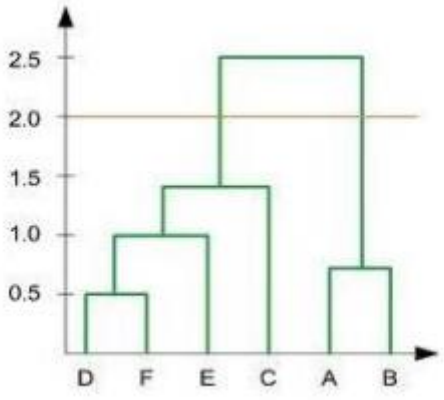
MACHINE LEARNING

(Assignment 3)

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is an application of clustering?
   Ans->b. Market trend prediction
2. On which data type, we cannot perform cluster analysis?
   Ans-> d. None
3. Netflix's movie recommendation system uses
   Ans-> b. Unsupervised learning
4. The final output of Hierarchical clustering is
   Ans -> a. The number of cluster centroids
5. Which of the step is not required for K-means clustering?
   Ans -> d. None
6. Which is the following is wrong?
   Ans -> c. k-nearest neighbour is same as k-means
7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?
   i. Single-link
   ii. Complete-link
   iii. Average-link
   Ans-> d. 1, 2 and 3
8. Which of the following are true?
   i. Clustering analysis is negatively affected by multicollinearity of features
   ii. Clustering analysis is negatively affected by heteroscedasticity
   Ans -> A. 1 only
9. In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?



Ans-> b. 2

10. For which of the following tasks might clustering be a suitable approach?

    Ans -> a.Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

11. Given, six points with the following attributes:

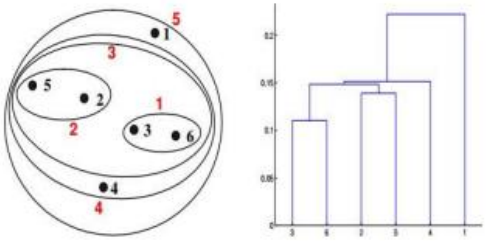
point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table : X-Y coordinates of six points.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:



Ans -> a)

12. Given, six points with the following attributes:

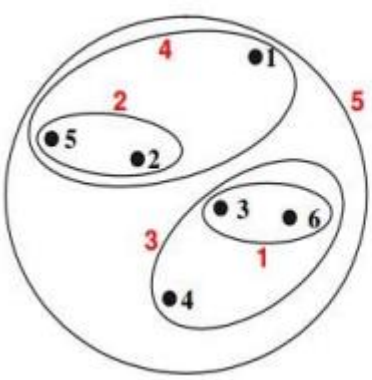
point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

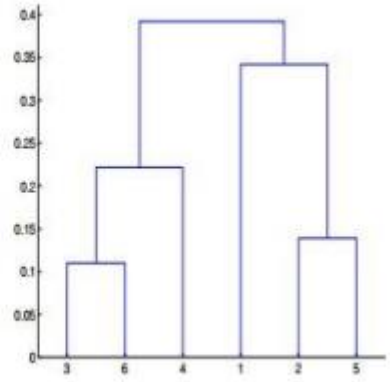
Table : X-Y coordinates of six points.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.Ans –





Ans-> B.

Q13 to Q14 are subjective answers type questions, Answers them in their own words briefly

13. What is the importance of clustering?

Ans- > Before knowing the Importance of Clustering . Lets Understand what is Cluster.

Cluster -> In this technique , Which groups the unlablled dataset .
A way of grouping the data point into different clusters consisting of similar data points.
The objects with possible similarities remain in a group that has less or no similarities with another group.

➔ **It is used by Amazon , Netflix , Youtube , Instagram in its recommendation system provide . Recommendation as per the past search of the products .**

Lets Understand Importance of Clustering

We all use Youtube for wathing videos and shorts . But all person likes different different creators . Like , Someone line Cooking Videos , Someone Like Comedy Videos , and Someone like Web Series .

So basically what Clustering Does . We all saw in real life which videos we like to watch them in youtube they show only that type of videos . How it can Know what you like ? So here we extract the data of few people as per their search and watch history . we start create group by similar search . and make make a model by using unsupervised machine learning algorthims . and make cluster as per data .. and then we go for production .

So after we do the clustering we see that . Peoples get attached with that particular Application because people this youtube is very smart what we like they show us .

- ➔ **It gives good impression to user .**
- ➔ **Helps in understanding the natural grouping in a dataset**
- ➔ **Organizing data into clusters shows internal structure of the data**
- ➔ **Techniques for clustering is useful in knowledge discovery in data**

14. How can I improve my clustering performance?

Ans ->