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?
   a. Biological network analysis
   b. Market trend prediction
   c. Topic modeling
   d. All of the above

Ans - All of the above(d)

2. On which data type, we cannot perform cluster analysis?
   a. Time series data
   b. Text data
   c. Multimedia data
   d. None

Ans - None(d)

3. Netflix's movie recommendation system uses-
   a. Supervised learning
   b. Unsupervised learning
   c. Reinforcement learning and Unsupervised learning
   d. All of the above

Ans - Supervised learning(a)

4. The final output of Hierarchical clustering is-
   a. The number of cluster centroids
   b. The tree representing how close the data points are to each other
   c. A map defining the similar data points into individual groups
   d. All of the above

Ans - A map defining the similar data points into individual groups(c)

5. Which of the step is not required for K-means clustering?
   a. A distance metric
   b. Initial number of clusters
   c. Initial guess as to cluster centroids
   d. None

Ans - A distance metric(a)

6. Which is the following is wrong?
   a. k-means clustering is a vector quantization method
   b. k-means clustering tries to group n observations into k clusters
   c. k-nearest neighbour is same as k-means
   d. None

Ans - 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

Options:

a. 1 and 2
b. 1 and 3
c. 2 and 3
d. 1, 2 and 3

Ans - 1, 2 and 3 (d)

8. Which of the following are true?

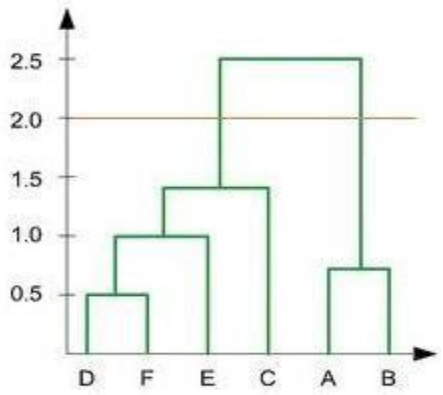
i. Clustering analysis is negatively affected by multicollinearity of features
ii. Clustering analysis is negatively affected by heteroscedasticity

Options:

a. 1 only
b. 2 only
c. 1 and 2
d. None of them

Ans - 1 only(a)

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?



a. 2
b. 4
c. 3
d. 5

Ans - 2(a)

10. For which of the following tasks might clustering be a suitable approach?
    a. Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.
    b. Given a database of information about your users, automatically group them into different market segments.
    c. Predicting whether stock price of a company will increase tomorrow.
    d. Given historical weather records, predict if tomorrow's weather will be sunny or rainy.

Ans - Given a database of information about your users, automatically group them into different market segments. (b)

11. Given, six points with the following attributes:

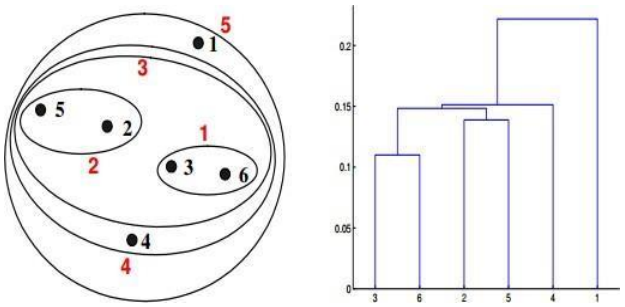
point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

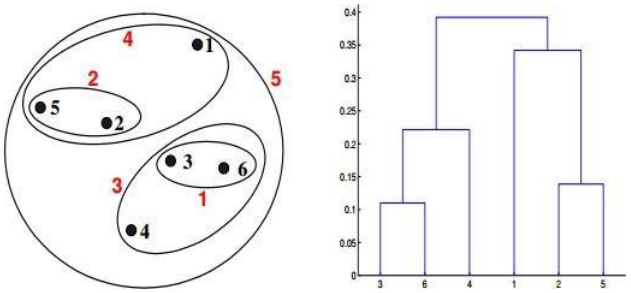
Table : X-Y coordinates of six points.

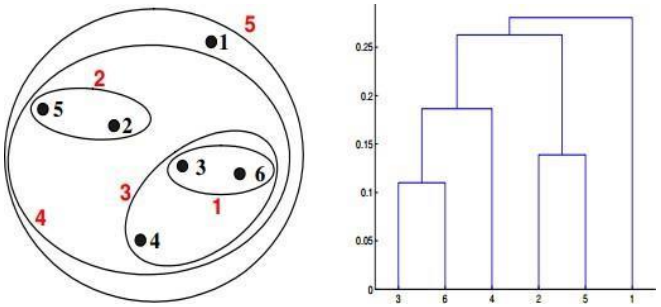
	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

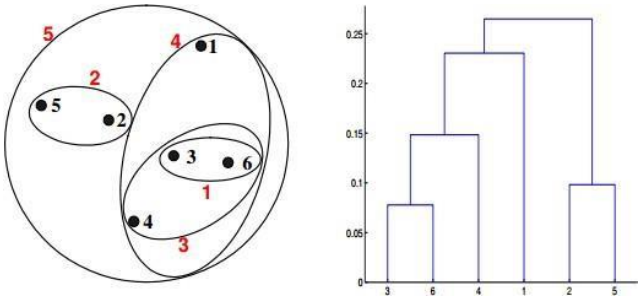
Table : Distance Matrix for Six Points

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

a. 

b. 

c. 

d. 

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

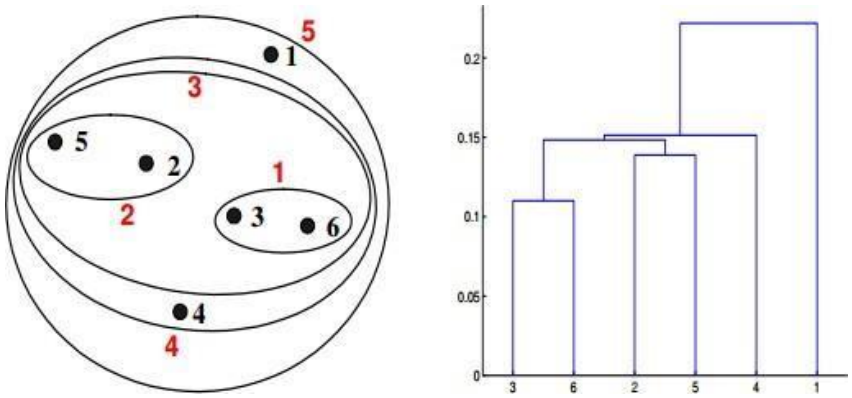
Table : X-Y coordinates of six points.

	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

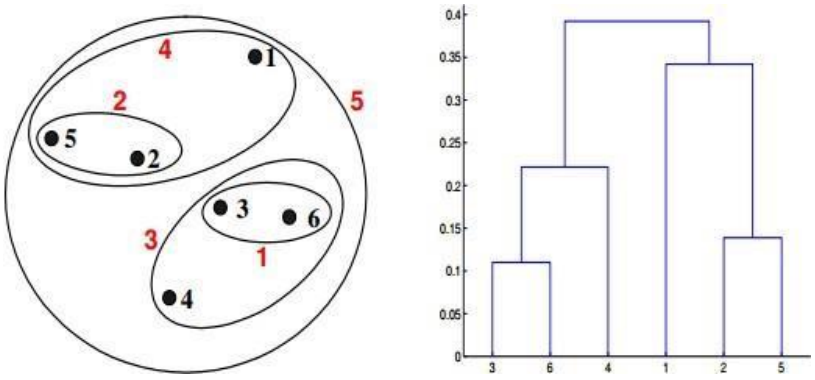
Table : Distance Matrix for Six Points

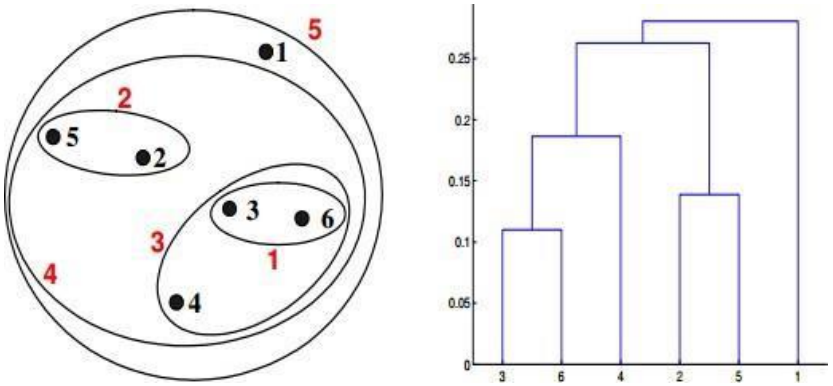
Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.

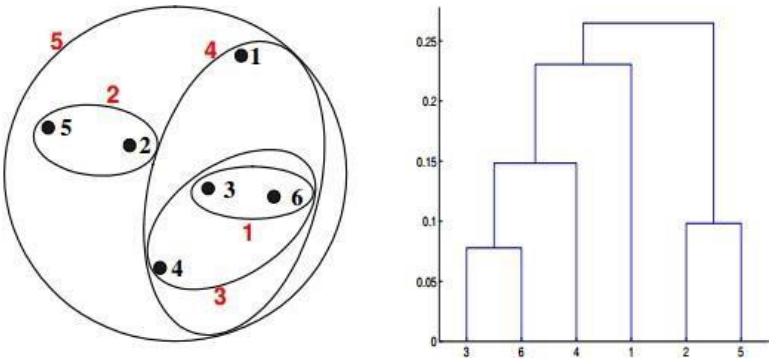
a.



b.



c. 

d. 

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 - Utilizing clustering techniques allows one to restart the local search process and eliminate inefficiencies. Clustering also helps in determining the inner structure of the data. The model analysis and vector region of attraction have both employed this clustering technique. Understanding natural groupings in a dataset is made easier by clustering. Their goal is to make it logical to separate the data into different logical divisions. The methodologies used and the discovery of hidden patterns influence clustering quality. They have a wide range of uses, including image processing, spatial research, weblog similarity metrics, marketing economic study, and more. They help in the detection of credit card fraud through outlier detections.

14. How can I improve my clustering performance?

Ans - Using ICA blind source separation during in the graph Laplacian embedding phase can easily boost the performance of graph-based clustering. Clustering performance is enhanced by employing either RICA or SFT to apply unsupervised feature learning to input data.