MACHINE LEARNING WORKSHEET 3

Q1. 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: (d) All of the above

Q2. On which data type, we cannot perform cluster analysis?

a. Time series data

b. Text data

c. Multimedia data

d. None

ANS: (d) None

Q3. Netflix's movie recommendation system uses

a. Supervised learning

b. Unsupervised learning

c. Reinforcement learning and Unsupervised learning

d. All of the above

ANS: (c) Reinforcement learning and Unsupervised learning

Q4. 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: (b) The tree representing how close the data points are to each other

Q5. 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:(d) None

Q6. 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:(c) k-nearest neighbour is same as k-means

Q7. 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**:(d)1, 2 and 3**

Q8. 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
b. 2 only
c. 1 and 2
d. None of them

ANS: (a)1 only b

Q9. In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?

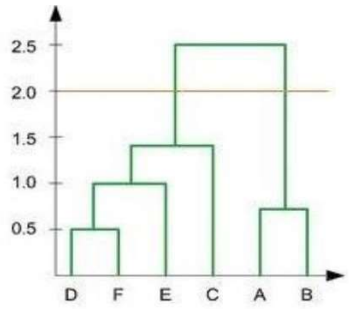
a. 2

b. 4

c. 3

d. 5

ANS:(a) 2



Q10. 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:(a) Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

Q11. 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

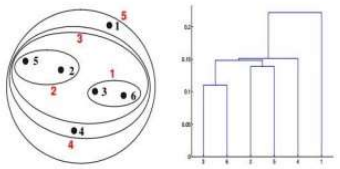
Table : X-Y coordinates of six points.

	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

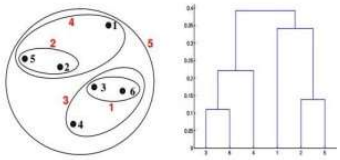
Table : Distance Matrix for Six Points

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

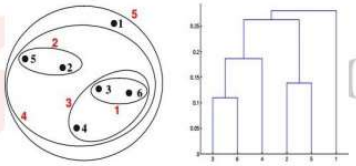
a.



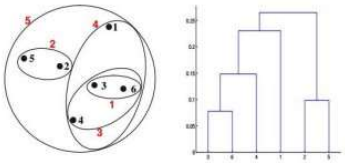
b.



c.



d.



ANS:(a)

Q12. 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

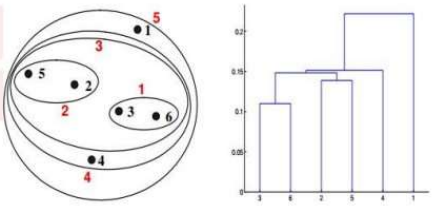
Table : X-Y coordinates of six points.

	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

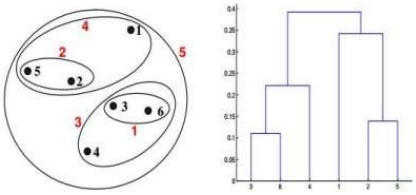
Table : Distance Matrix for Six Points

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

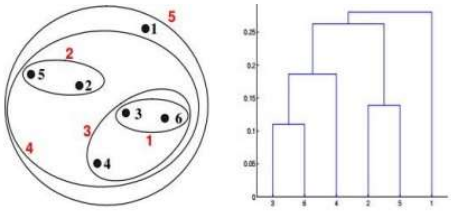
a.



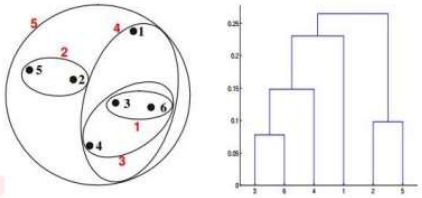
b.



c.



d.



ANS: **(b)**

Q13. What is the importance of clustering?

ANS: **a.** Having clustering methods helps in restarting the local search procedure and remove the inefficiency. In addition, clustering helps to determine the internal structure of the data.

b. This clustering analysis has been used for model analysis, vector region of attraction.

c. Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings.

d. Clustering quality depends on the methods and the identification of hidden patterns.

e. They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.

Q14. How can I improve my clustering performance?

ANS: Clustering is one example of how machine learning can help organizations improve processes by using meaningful data to inform changes. While embracing machine learning can be an overwhelming process for an organization, starting with a small pilot project can help pave the way for inspiring an entire organization to embrace the power of machine learning.