



Course Name: Introduction to Machine Learning

Assignment – Week 8 (Clustering)

TYPE OF QUESTION: MCQ/MSQ

Number of Question: 10

Total Marks: 10x2 = 20

Question 1:

Do the clustering results of the K-Means algorithm depend on the initial cluster centroid choices?

- A) Yes
- B) No

Correct Answer: A

Detailed Solution: K-Means clustering algorithm may converge on local minima which might also correspond to the global minima in some cases but not always. Different initial centroid choices may produce different clustering results.

Question 2:

Which of the following can act as possible termination conditions in K-Means?

- I. Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.
 - II. Centroids do not change between successive iterations.
- A) I only
 - B) II only
 - C) I and II

Correct Answer: C

Detailed Solution: Both the conditions can act as possible termination conditions.



Question 3:

Assume, you want to cluster 7 observations into 3 clusters using K-Means clustering algorithm. After first iteration the clusters: C1, C2, C3 has the following observations:

C1: {(1,1), (4,4), (7,7)}

C2: {(0,4), (4,0)}

C3: {(5,5), (9,9)}

What will be the cluster centroids after the first iteration?

- A) C1: (4,4), C2: (2,2), C3: (7,7)
- B) C1: (2,2), C2: (0,0), C3: (5,5)
- C) C1: (6,6), C2: (4,4), C3: (9,9)
- D) None of these

Correct Answer: A

Detailed Solution:

Finding centroid for data points in cluster C1 = $((1+4+7)/3, (1+4+7)/3) = (4, 4)$

Finding centroid for data points in cluster C2 = $((0+4)/2, (4+0)/2) = (2, 2)$

Finding centroid for data points in cluster C3 = $((5+9)/2, (5+9)/2) = (7, 7)$

Hence, C1: (4,4), C2: (2,2), C3: (7,7)

Question 4:

In single-link clustering, the similarity of two clusters is the similarity of their most similar members. What is the time complexity of the single-link clustering algorithm? (Note: n is the number of data points)

- A) $O(n^2)$
- B) $O(n^2 \log n)$
- C) $O(n^3 \log n)$
- D) $O(n^3)$

Correct Answer. A

Detailed Solution: Refer to the lecture.



Question 5:

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

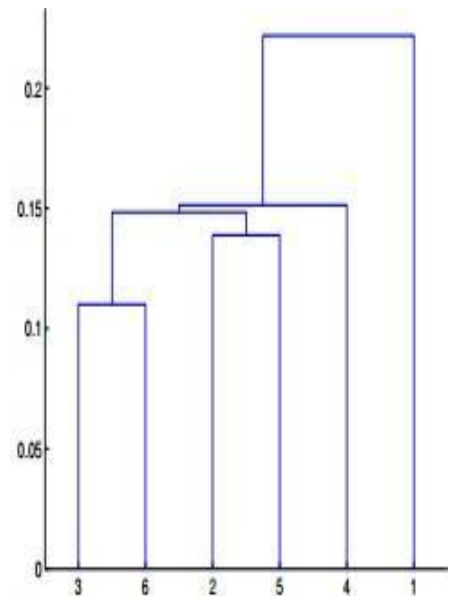
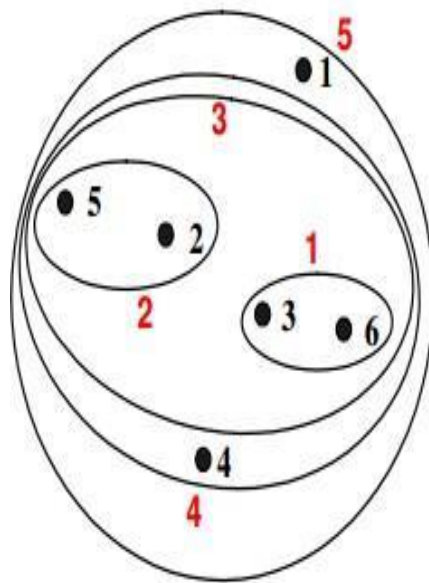
Table 1: x-y coordinates of six points

	p1	p2	p3	p4	p5	p6
p1	0.000	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	0.000	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

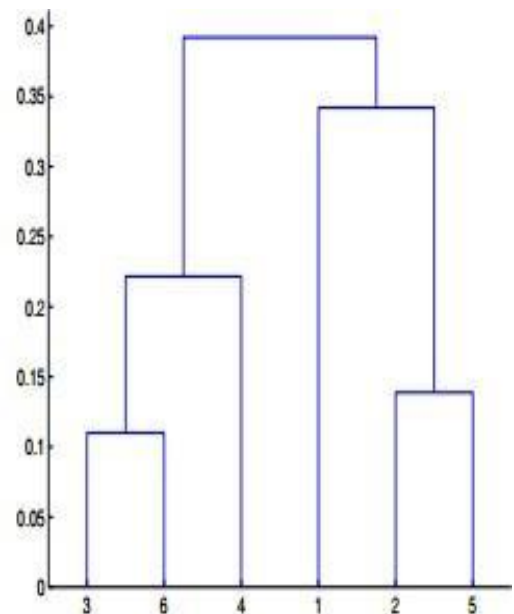
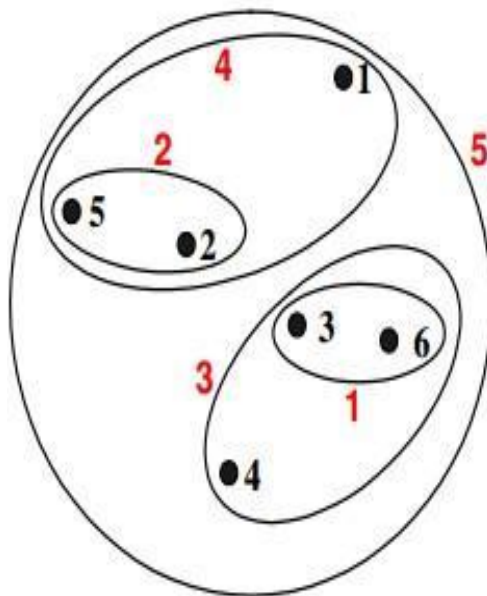
Table 2: Distance Matrix for six points

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

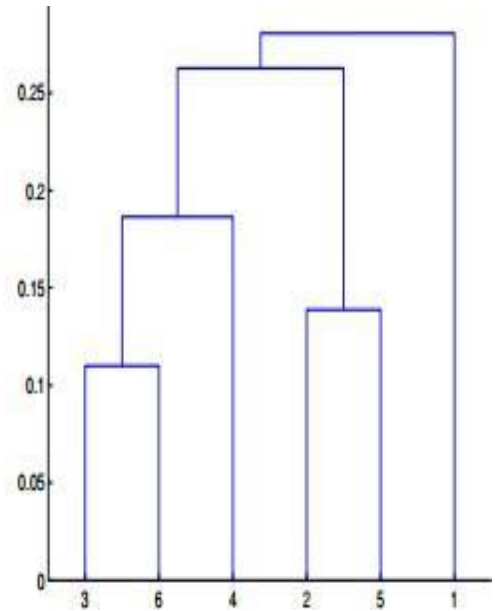
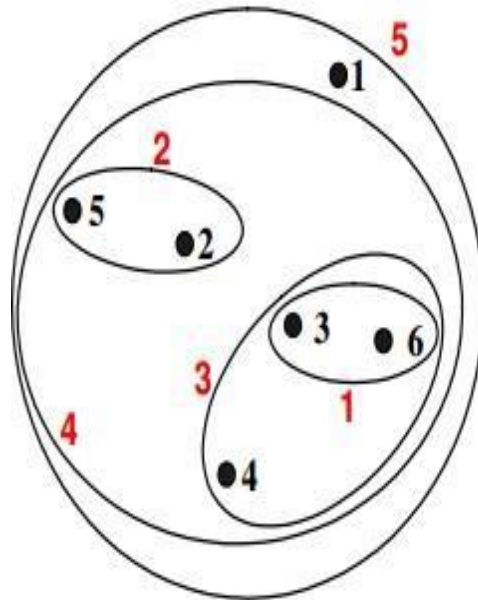
A)



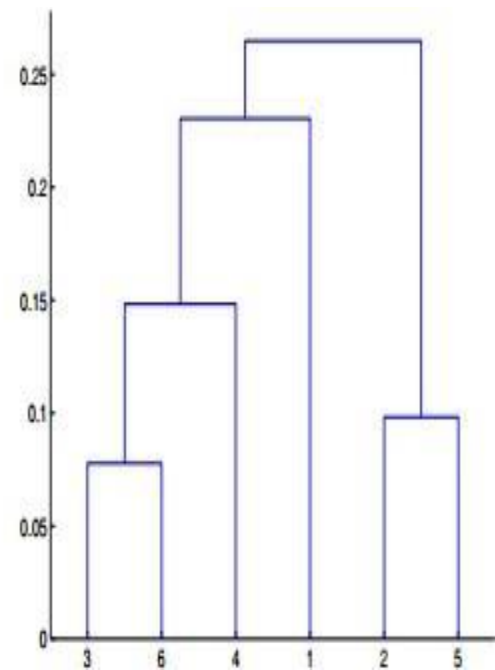
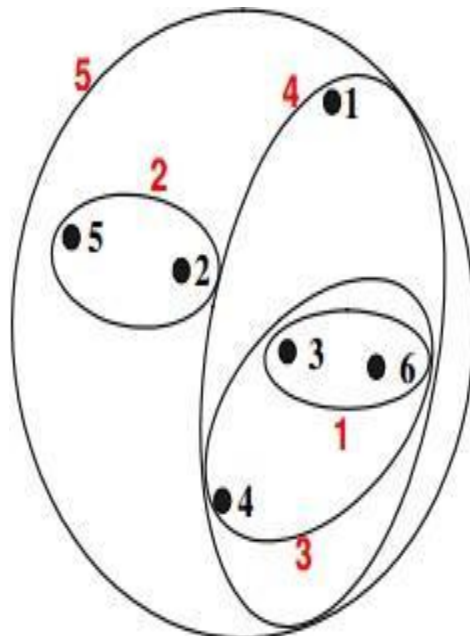
B)



C)



D)





Correct Answer: A

Detailed Solution: For the single link or MIN version of hierarchical clustering, the proximity of two clusters is defined to be the minimum of the distance between any two points in the different clusters. For instance, from the table, we see that the distance between points 3 and 6 is 0.11, and that is the height at which they are joined into one cluster in the dendrogram. As another example, the distance between clusters $\{3, 6\}$ and $\{2, 5\}$ is given by $\text{dist}(\{3, 6\}, \{2, 5\}) = \min(\text{dis}(3, 2), \text{dist}(6, 2), \text{dist}(3, 5), \text{dist}(6, 5)) = \min(0.1483, 0.2540, 0.2843, 0.3921) = 0.1483$.

Question 6:

Is it possible that assignment of observations to clusters does not change between successive iterations of K-means?

- A) Yes
- B) No
- C) Can't say
- D) None of these

Correct Answer: A

Detailed Solution: When the K-means has reached the global or local minima, it will not alter the assignment of data points to clusters in successive iterations.

Question 7:

Which of the following is not a clustering approach?

- A) Hierarchical
- B) Partitioning
- C) Bagging
- D) Density-Based

Correct Answer: C

Detailed Solution: Bagging is not a clustering technique.

Question 8:

In which of the following cases will K-Means clustering fail to give good results?

- A) Data points with outliers
- B) Data points with round shapes
- C) Data points with non-convex shapes
- D) Data points with different densities



Correct Answer: A, C, D

Detailed Solution: K-Means clustering algorithm fails to give good results when the data contains outliers, the density spread of data points across the data space is different and the data points follow non-convex shapes.

Question 9:

Given, $A = \{0,1,2,5,6\}$ and $B = \{0,2,3,4,5,7,9\}$, calculate Jaccard Index of these two sets.

- A) 0.50
- B) 0.25
- C) 0.33
- D) 0.41

Correct Answer. C

Detailed Solution: Jaccard Index $J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{3}{9} = 0.33$

Question 10:

Which of the following statements is/are not true about k-means clustering?

- A) It is an unsupervised learning algorithm
- B) Overlapping of clusters is allowed in k-means clustering
- C) It is a hard-clustering technique
- D) k is a hyperparameter in k-means

Correct Answer: B

Detailed Solution: Overlapping of clusters is not allowed in k-means.

END