

## ASSIGNMENT – 1

### Machine learning

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

1. What is the most appropriate no. of clusters for the data points represented by the following dendrogram:

- a) 2
- b) 4**
- c) 6
- d) 8

**Ans: b**

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

- 1. Data points with outliers
- 2. Data points with different densities
- 3. Data points with round shapes
- 4. Data points with non-convex shapes

Options:

- a) 1 and 2
- b) 2 and 3
- c) 2 and 4
- d) 1, 2 and 4**

**Ans:d**

3. The most important part of is selecting the variables on which clustering is based.

- a) interpreting and profiling clusters
- b) selecting a clustering procedure
- c) assessing the validity of clustering
- d) formulating the clustering problem**

**Ans:d**

4. The most commonly used measure of similarity is the or its square.

- a) Euclidean distance**
- b) city-block distance
- c) Chebyshev's distance
- d) Manhattan distance

**Ans:a**

5.

is a clustering procedure where all objects start out in one giant cluster. Clusters are formed by dividing this cluster into smaller and smaller clusters.

a) Non-hierarchical clustering

**b) Divisive clustering**

c) Agglomerative clustering

d) K-means clustering

**Ans:b**

6. Which of the following is required by K-means clustering?

a) Defined distance metric

b) Number of clusters

c) Initial guess as to cluster centroids

**d) All answers are correct**

**Ans:d**

7. The goal of clustering is to

**a) Divide the data points into groups**

b) Classify the data point into different classes

c) Predict the output values of input data points

d) All of the above

**Ans:a**

8. Clustering is a

a) Supervised learning

**b) Unsupervised learning**

c) Reinforcement learning

d) None

**Ans:b**

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

a) K- Means clustering

b) Hierarchical clustering

c) Diverse clustering

**d) All of the above**

**Ans:d**

10. Which version of the clustering algorithm is most sensitive to outliers?

**a) K-means clustering algorithm**

b) K-modes clustering algorithm

c) K-medians clustering algorithm

d) None

**Ans:a**

11. Which of the following is a bad characteristic of a dataset for clustering analysis

- a) Data points with outliers
- b) Data points with different densities
- c) Data points with non-convex shapes
- d) **All of the above**

**Ans:d**

12. For clustering, we do not require

- a) **Labeled data**
- b) Unlabeled data
- c) Numerical data
- d) Categorical data

**Ans:a**

**Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly.**

**13. How is cluster analysis calculated?**

Cluster analysis is an exploratory data analysis that tries to identify structures within the data. It tries to identify homogeneous group of cases if the grouping is not previously known. because it is exploratory, it does not make any distinction between dependent and independent variables.

It is also called as segmentation analysis or taxonomy analysis.

**14. How is cluster quality measured?**

To measure the quality of clustering we can use the average silhouette coefficient value of all objects in the dataset.

**15. What is cluster analysis and its types?**

Cluster analysis is a task of grouping a set of data points in such a way that they can be characterized by their relevancy to one another.

The types are Centroid clustering, Density Clustering, Distribution Clustering and Connectivity Clustering.