

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?

d. All of the above

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

d. None

3. Netflix's movie recommendation system uses.

c. Reinforcement learning

4. The final output of Hierarchical clustering is

b. The tree representing how close the data points are to each other

5. Which of the step is not required for K-means clustering?

d. None

6. Which of the following is wrong?

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 Options:

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

Options:

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?

a. 2

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

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

11. Given, six points with the following attributes: Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

A

12. Given, six points with the following attributes: Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.

B

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

13. What is the importance of clustering?

Clustering is important in data analysis and data mining applications. It is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups (clusters).

14. How can I improve my clustering performance?

Using the new clustering pipeline presented here, effective clustering performance can be obtained without employing deep clustering algorithms and their accompanying hyper-parameter tuning procedure.