

Machine learning assignment 3

1. Which of the following is an application of clustering?

Ans : d. All of the above

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

Ans : d. None

3. Netflix's movie recommendation system uses

Ans : c. Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering is

Ans : 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?

Ans : d. None

6. Which is the following is wrong?

Ans : 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

Ans : 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

Ans : 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?

Ans : a. 2

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

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

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

Ans : a

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

Ans : b

13. What is the importance of clustering?

Ans : Clustering is a significant component of Machine Learning. It has a wide range of applications and advantages in real-world scenarios. It is widely used by many companies like LinkedIn, Amazon, and Netflix to find identify similar personalities/customers or to find similar products liked by the customers. Based on this they recommend the products for the customers to buy or suggest people get connected with.

Few applications of clustering are :

- Customer segmentation where similar customers are identified based on their purchase behaviour and are grouped for further discounts or offers
- Image segmentation groups the similar attributes of pixels in an image to make it easier to analyze.
- In outlier detection, the data points that lie in a high-density area is considered to be normal data point whereas the data points that are away from all other data points and does not belong to any group are considered outliers
- Data analysis: when it comes to analyzing the data it is very helpful in getting insights into the data by explaining the characteristics of each group/cluster.

14. How can I improve my clustering performance?

Ans : The basic idea is that you set the maximum variance you allow in each cluster. You start with as many clusters as data points and then you evolve clusters by

- merging neighbouring clusters if the resulting cluster's variance is below the threshold
- isolating elements that are far if a cluster's variance is above the threshold
- or moving some elements between neighbouring clusters if it decreases the sum of squared errors