

MACHINE LEARNING

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

1. d. All of the above
2. d. None
3. c. Reinforcement learning and Unsupervised learning
4. b. The tree representing how close the data points are to each other
5. d. None
6. c. k-nearest neighbor is same as k-means
7. d. 1, 2 and 3
8. a. 1 only
9. a. 2
10. b. Given a database of information about your users, automatically group them into different market segments.
11. a
12. b

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

13. What is the importance of clustering?

Clustering in machine learning is an essential component and makes life so much easier in creating new machine learning methods. It mainly divides many unstructured data sets into clusters and, according to the common attributes present in them, it helps create more and more clusters.

The purpose of clustering and classification algorithms is to make sense of and extract value from large sets of structured and unstructured data. If you're working with huge volumes of unstructured data, it only makes sense to try to partition the data into some sort of logical groupings before attempting to analyze it.

Clustering is useful for exploring data. If there are many cases and no obvious groupings, clustering algorithms can be used to find natural groupings. Clustering can also serve as data pre-processing step to identify homogeneous groups on which to build supervised models.

Applications of Clustering in different fields

Marketing: It can be used to characterize & discover customer segments for marketing purposes.

Biology: It can be used for classification among different species of plants and animals.

Libraries: It is used in clustering different books on the basis of topics and information.

Insurance: It is used to acknowledge the customers, their policies and identifying the frauds.

City Planning: It is used to make groups of houses and to study their values based on their geographical locations and other factors present.

Earthquake studies: By learning the earthquake-affected areas we can determine the dangerous zones.

14. How can I improve my clustering performance?

Clustering is an unsupervised machine learning methodology that aims to partition data into distinct groups, or clusters. There are a few different forms including hierarchical, density, and similarity based. Each have a few different algorithms associated with it as well.

K-means clustering algorithm can be significantly improved by using a better initialization technique, and by re starting the algorithm, or avoiding unbalanced cluster size. When the data has overlapping clusters K - means can improve the results of the initialization technique. When the data has well separated clusters, the performance of k-means depends completely on the goodness of the initialization. Initialization using simple furthest point heuristic (Maxmin) reduces the clustering error of k-means from 15% to 6%, on average.