Great Learning
POWER AHEAD

Unsupervised Supervised Learning Interview Questions

1. If you want to find different groups of customers based on their salary and purchase patterns, which kind of machine learning algorithm would you use?
2. If we have data about employees' experience in years, education, and salary and we want to identify employees that need to undergo professional training, how can we use clustering to accomplish this?
3. In general, when do we use unsupervised learning?
4. How is KNN different from k-means clustering?
5. What is the minimum number of variables/features required to do clustering?
6. Can we apply clustering for multidimensional data?
7. What are the pros and cons of Hierarchical clustering?
8. What is an agglomerative method in hierarchical clustering?
9. Can a decision tree be used to form a cluster?
10. Explain Principal Component Analysis? Explain what it means that the first principal component explains 10% of the variation.
11. What is the role of Inertia In K Means Clustering?
12. Explain the working of dendrogram in the agglomerative Hierarchical Clustering Algorithm.
13. How the K-Value is decided in K Means Clustering. Is the domain knowledge or pattern identified in the data or both?
14. Why is the Euclidean distance most preferable instead of the Manhattan distance for K-means?
15. How do you evaluate a Good Cluster vs a Bad One? What all metrics are involved in checking the quality of clusters formed.

16. How PCA reduces the Dimensions of the Data from n - Dimensions to 2-Dimensions. Does it have to do something with 2D Planes or 3D Planes?

17. Why do we need dimensionality reduction? What are its drawbacks?

18. Can PCA be used to reduce the dimensionality of a highly nonlinear dataset?

19. Is rotation necessary in PCA? If yes, Why? What will happen if you don't rotate the components?

20. How can you evaluate the performance of a dimensionality reduction algorithm on your dataset?

21. What are some Stopping Criteria for k-Means Clustering?

22. Explain some cases where k-Means clustering fails to give good results

23. How do Forgy Initialization and Random Partition Initialization compare with each other.

24. How does the Curse of Dimensionality affect k-Means Clustering?

25. How to get dissimilarity percentage for clustering in Hierarchical cluster analysis?

26. . How is the parameter "Distance-function" estimated in the DBSCAN Algorithm?

27. Why does there arise a need for DBSCAN when we already have other clustering Algorithms?

28. What are some different types of clustering structures that are used in clustering algorithms?

29. How do the cluster algorithms work on detecting anomalies when the cluster size is different?

30. Compare between the k-means and PCA?