1. What is one of the main differences between supervised and unsupervised learning?

a) Supervised learning requires labeled data, while unsupervised learning does not need labeled data.
b) Unsupervised learning requires labeled data, while supervised learning does not need labeled data.
c) Supervised learning always uses clustering techniques, while unsupervised learning does not.
d) Unsupervised learning always requires a pre-defined number of classes.

Answer: a) Supervised learning requires labeled data, while unsupervised learning does not need labeled data.
Explanation: Supervised learning relies on labeled data, where each input has an associated output label for the model to learn from. In contrast, unsupervised learning does not require labeled data and often focuses on finding patterns or structures in the input data without explicit guidance.

2. Which criterion function is commonly used in K-means clustering to minimize the intra-cluster variance?

a) Entropy
b) Gini impurity
c) Davies-Bouldin index
d) Sum of squared distances

Answer: d) Sum of squared distances
Explanation: K-means clustering aims to minimize the sum of squared distances between

data points and their corresponding cluster centroids. This criterion function helps in optimizing the placement of centroids to form clusters with minimal intra-cluster variance.

3. Which clustering technique involves iteratively reassigning data points to clusters based on the closest centroid?

a) Hierarchical clustering
b) DBSCAN
c) K-means clustering
d) Gaussian Mixture Models

Answer: c) K-means clustering
Explanation: K-means clustering iteratively assigns data points to the nearest centroid and updates the centroids based on the mean of the assigned points. This process continues until convergence, resulting in clusters with minimized intra-cluster variance.

4. What is the primary objective of cluster validation in unsupervised learning?

a) To determine the number of clusters present in the data
b) To ensure that every data point is assigned to a cluster
c) To maximize inter-cluster variance
d) To minimize intra-cluster variance

Answer: a) To determine the number of clusters present in the data
Explanation: Cluster validation techniques aim to assess the quality of clustering results and determine the optimal number of clusters present in the data. This helps in evaluating the effectiveness of clustering algorithms and selecting the appropriate number of clusters for further analysis.

5. Which of the following is NOT a common hierarchical clustering method?

a) Single-linkage clustering
b) Complete-linkage clustering
c) Average-linkage clustering
d) Iterative-linkage clustering

Answer: d) Iterative-linkage clustering
Explanation: Iterative-linkage clustering is not a common hierarchical clustering method. Instead, single-linkage, complete-linkage, and average-linkage clustering are commonly used methods to construct hierarchical cluster trees based on the distance between data points or clusters.

6. Which cluster validation index assesses the compactness and separation between clusters?

a) Silhouette score
b) Davies-Bouldin index
c) Calinski-Harabasz index
d) Dunn index

Answer: a) Silhouette score
Explanation: The silhouette score measures the compactness and separation between clusters by computing the mean silhouette coefficient for all data points. A higher silhouette score indicates better-defined clusters with greater inter-cluster separation and intra-cluster cohesion.

7. In hierarchical clustering, what does the “linkage” criterion refer to?

a) The way clusters are formed
b) The distance metric used to measure dissimilarity between clusters
c) The process of assigning data points to clusters
d) The number of clusters in the final output

Answer: b) The distance metric used to measure dissimilarity between clusters
Explanation: In hierarchical clustering, the “linkage” criterion refers to the method used to calculate the distance or dissimilarity between clusters. Common linkage criteria include single-linkage, complete-linkage, and average-linkage, which determine how clusters are merged during the clustering process.

8. What is the primary drawback of the K-means clustering algorithm?

a) It requires a pre-defined number of clusters
b) It is sensitive to the initial choice of cluster centroids
c) It cannot handle high-dimensional data
d) It always produces spherical clusters

Answer: b) It is sensitive to the initial choice of cluster centroids
Explanation: K-means clustering is sensitive to the initial placement of cluster centroids, which can result in different clustering outcomes based on the initial centroids’ locations. This sensitivity makes the algorithm prone to converging to local optima rather than the global optimum.

9. Which unsupervised learning technique is most suitable for identifying outliers and dense regions in high-dimensional data?

a) K-means clustering

b) DBSCAN
c) Hierarchical clustering
d) Gaussian Mixture Models

Answer: b) DBSCAN
Explanation: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is well-suited for identifying outliers and dense regions in high-dimensional data. It does not require a pre-defined number of clusters and can handle irregularly shaped clusters by defining clusters based on density connectivity.

10. What is a potential limitation of hierarchical clustering compared to partitional clustering algorithms like K-means?

a) Hierarchical clustering requires a pre-defined number of clusters
b) Hierarchical clustering is computationally less efficient
c) Hierarchical clustering cannot handle non-linearly separable data
d) Hierarchical clustering can be challenging to interpret for large datasets

Answer: b) Hierarchical clustering is computationally less efficient
Explanation: Hierarchical clustering algorithms, especially agglomerative hierarchical clustering, are computationally less efficient compared to partitional clustering algorithms like K-means. This is because hierarchical clustering involves constructing and maintaining a tree-like structure of clusters, which can be computationally intensive for large datasets.

Related posts:

1. Introduction to Information Security
2. Introduction to Information Security MCQ

3. Introduction to Information Security MCQ
4. Symmetric Key Cryptography MCQ
5. Asymmetric Key Cryptography MCQ
6. Authentication & Integrity MCQ
7. E-mail, IP and Web Security MCQ