Quiz 4

Score: 13/14

1. What is hierarchical clustering?

- A method for classifying objects into a structure of clusters ✓
- A method for dimensionality reduction
- A method for outlier detection
- A method for regression analysis

Explanation

Hierarchical clustering is a type of clustering algorithm that builds a hierarchy of clusters. It can be agglomerative (bottom-up) or divisive (top-down).

2. Which of the following is a linkage criterion used in hierarchical clustering?

- Single-linkage ✓
- Spectral clustering
- Random noise clustering
- DBSCAN clustering

Explanation

Linkage criteria determine the distance between clusters and impact the structure of the dendrogram in hierarchical clustering. Common linkage criteria include single-linkage, complete-linkage, average-linkage, and ward's method.

3. In DBSCAN clustering, what is the meaning of 'epsilon'?

- Minimum number of points to form a cluster
- Maximum distance between two data points to form a cluster
- Maximum distance between two data points to be considered as in the same neighborhood ✓
- Minimum distance to form a core object

Explanation

'Epsilon' is the maximum distance between two data points to be considered as in the same neighborhood. It is a crucial parameter in DBSCAN algorithm.

4. What does the 'minPts' parameter represent in DBSCAN clustering?

- Maximum distance between two data points to form a cluster

Explanation

Minimum distance to form a core object

Minimum number of points to form a cluster

Maximum number of iterations in the algorithm

'minPts' defines the minimum number of data points required to form a dense region in DBSCAN clustering. It influences the differentiation of core, border, and outlier points in the dataset.

5. What is the time complexity of hierarchical clustering?

O(n^2)

O(n log n)

O(n^3)

O(n)

Explanation

The time complexity of hierarchical clustering is generally O(n^3), where n is the number of data points. This makes it computationally expensive for large datasets.

6. Which of the following is a disadvantage of hierarchical clustering?

Can handle large datasets efficiently

Not sensitive to noise and outliers

Computationally expensive for large datasets

Requires predetermined number of clusters

Explanation

One disadvantage of hierarchical clustering is its computational complexity, especially for large datasets. It also suffers from sensitivity to noise and outliers.

7. In hierarchical clustering, what do dendrograms represent?

Visual representation of k-means clusters

Graphical representation of DBSCAN clustering

Illustration of the arrangement of clusters

Visualization of support vector machine decisions

Explanation

Dendrograms are tree-like diagrams that illustrate the arrangement of the clusters produced by hierarchical clustering. They are useful for understanding the hierarchical relationships among data points.

8. **Which type of clustering algorithm is DBSCAN?**

- Centroid-based clustering
- Density-based clustering
- Partitioning clustering
- Agglomerative clustering

Explanation

DBSCAN is a density-based clustering algorithm that can identify clusters of irregular shapes and is robust to noise and outliers in the data.

9. **Which of the following is a characteristic of DBSCAN clustering?**

- Suitable only for spherical clusters
- Sensitive to noise and outliers
- Can identify clusters of arbitrary shapes and sizes
- Requires predefined number of clusters

Explanation

DBSCAN can identify clusters of arbitrary shapes and sizes, making it effective in scenarios where clusters have varying densities and shapes.

10. **In DBSCAN clustering, what is a 'core point'?**

- A noise point
- A border point
- A data point with at least 'minPts' data points within 'epsilon'
- A point with maximum distance to form a cluster

Explanation

A 'core point' is a data point that has at least 'minPts' data points within a distance of 'epsilon', indicating that it is in a dense region and likely part of a cluster.

11. **What measure is commonly used to evaluate the quality of clustering in hierarchical clustering?**

- Silhouette coefficient
- Davies-Bouldin index
- Cophenetic correlation coefficient
- Calinski-Harabasz index

Explanation

The most common measure for evaluating the quality of clustering in hierarchical clustering is the cophenetic correlation coefficient, which quantifies how faithfully a dendrogram preserves the pairwise distances between the original data points.

12. **Which of the following distance measures is commonly used in hierarchical clustering?**

- Jaccard distance
- Cosine similarity
- Euclidean distance
- Hamming distance

Explanation

In hierarchical clustering, commonly used distance measures include Euclidean distance, Manhattan distance, and Mahalanobis distance, among others.

13. **What is the main advantage of hierarchical clustering over k-means clustering?**

- Faster convergence
- Scalability to large datasets
- Does not require the number of clusters to be specified
- Robustness to noise and outliers

Explanation

One of the main advantages of hierarchical clustering over k-means clustering is that it does not require the number of clusters to be specified a priori, allowing for a more exploratory data analysis approach.

14. **In DBSCAN clustering, what is the purpose of the 'eps-neighborhood' of a point?**

- To identify noise points
- To define the maximum cluster size
- To determine the density of the point
- To identify core points and expand clusters

Explanation

The 'eps-neighborhood' of a point includes all points within a distance of 'epsilon' from the given point. It is used to identify core points and expand clusters in DBSCAN.