Quiz 3 sur le Score de Silhouette

1. Qu'est-ce que le score de silhouette mesure principalement ?

A. La cohésion interne des clusters et la séparation des clusters voisins.

B. La vitesse de convergence d'un algorithme de clustering.

C. La densité des clusters.

D. Le nombre de clusters dans les données.

2. Quelle valeur de score de silhouette indique des points bien groupés et bien séparés des autres clusters ?

A. -1

B. 0

C. 1

D. 0.5

3. Quelle valeur de score de silhouette indique que les points sont probablement mal assignés à un cluster ?

A. -1

B. 0

C. 1

D. 0.5

4. Que représente un score de silhouette proche de 0 ?

A. Les points sont bien groupés.

B. Les points sont à la frontière entre deux clusters.

C. Les points sont mal assignés.

D. Les points sont bien séparés des autres clusters.

5. Quel algorithme de clustering est souvent évalué en utilisant le score de silhouette ?

A. Régression linéaire

B. K-Means

C. Arbres de décision

D. Naive Bayes

6. Comment le score de silhouette est-il calculé pour chaque point de données ?

A. En mesurant la distance maximale entre le point et le centroïde du cluster.

B. En calculant la distance moyenne entre le point et tous les autres points du même cluster et la distance moyenne entre le point et tous les points des clusters voisins.

C. En utilisant la distance euclidienne entre tous les points.

D. En mesurant le nombre de points voisins.

7. Un score de silhouette moyen de 0.75 indique :

A. Les clusters sont mal formés.

B. Les clusters sont bien séparés et bien groupés.

C. Les points de données sont à la frontière entre deux clusters.

D. Les points de données sont mal assignés.

8. Pourquoi le score de silhouette est-il important dans l'évaluation des clusters ?

A. Il aide à déterminer le nombre optimal de clusters.

B. Il mesure la performance de l'algorithme en termes de temps d'exécution.

C. Il évalue la qualité de la séparation et de la cohésion des clusters.

D. Il compte le nombre total de points de données.

9. Quelle est la portée du score de silhouette ?

A. 0 à 1

B. -1 à 0

C. 0 à 10

D. -1 à 1

10. Qu'est-ce qu'une bonne pratique pour interpréter un score de silhouette ?

A. Un score de silhouette élevé signifie que les clusters sont bien définis et distincts.

B. Un score de silhouette élevé signifie que les points de données sont mal assignés.

C. Un score de silhouette élevé indique que l'algorithme a convergé rapidement.

D. Un score de silhouette élevé signifie qu'il y a beaucoup de bruit dans les données.

11. Quel type de données peut être utilisé pour calculer le score de silhouette ?

A. Données catégorielles uniquement.

B. Données numériques uniquement.

C. Données textuelles uniquement.

D. Données numériques et catégorielles après encodage.

12. Dans le contexte du score de silhouette, que signifie une séparation élevée ?

A. Les points sont proches du centre du cluster.

B. Les points sont bien séparés des autres clusters.

C. Les points sont équidistants de tous les autres points.

D. Les points sont proches les uns des autres dans un cluster.

13. Comment le score de silhouette peut-il être visualisé pour aider à évaluer la qualité du clustering ?

A. En utilisant un histogramme des distances.

B. En utilisant un graphique en barres.

C. En utilisant un diagramme en boîte.

D. En utilisant un diagramme des silhouettes.

14. Quelle est l'étape suivante après avoir calculé le score de silhouette pour un ensemble de données ?

A. Utiliser le score pour déterminer le nombre optimal de clusters.

B. Calculer la moyenne des distances euclidiennes.

C. Recalculer les centroids des clusters.

D. Appliquer une réduction de dimensionnalité.

15. Pourquoi le score de silhouette est-il préféré à l'indice de Rand ajusté pour évaluer les clusters ?

A. Parce qu'il est plus facile à calculer.

B. Parce qu'il ne nécessite pas de vérité terrain.

C. Parce qu'il est plus rapide.

D. Parce qu'il utilise la distance euclidienne.

16. Comment un score de silhouette peut-il être utilisé pour détecter des anomalies ?

A. Les points avec un score de silhouette proche de 1 sont des anomalies.

B. Les points avec un score de silhouette proche de 0 sont des anomalies.

C. Les points avec un score de silhouette négatif sont des anomalies.

D. Les points avec un score de silhouette positif sont des anomalies.

17. Quelle méthode de clustering n'utilise pas typiquement le score de silhouette pour évaluer la qualité des clusters ?

A. DBSCAN

B. K-Means

C. Agglomerative Clustering

D. Linear Regression

18. Que peut indiquer un score de silhouette moyen faible pour un ensemble de données ?

A. Les clusters sont bien formés.

B. Les points de données sont bien assignés.

C. Les clusters ne sont pas bien séparés ou les points sont mal assignés.

D. Les clusters sont trop denses.

19. Comment le score de silhouette est-il utile dans la validation croisée des modèles de clustering ?

A. Il mesure le temps d'exécution des modèles.

B. Il évalue la stabilité et la qualité des clusters sur différentes partitions de données.

C. Il ajuste automatiquement les hyperparamètres des modèles.

D. Il sélectionne les caractéristiques importantes pour le clustering.

20. Lorsqu'on utilise le score de silhouette pour choisir le nombre optimal de clusters, que cherche-t-on généralement ?

A. Le nombre de clusters qui minimise le score de silhouette.

B. Le nombre de clusters qui maximise le score de silhouette.

C. Le nombre de clusters qui donne un score de silhouette moyen.

D. Le nombre de clusters qui donne le score de silhouette le plus bas.