Quiz 4 sur l'Évaluation des Clusters (Intra-cluster et Inter-cluster)

1. Qu'est-ce que la cohésion intra-cluster mesure ?

A. La distance moyenne entre les points de différents clusters.

B. La distance moyenne entre les points et le centroïde du même cluster.

C. La variance totale des données.

D. La séparation entre les clusters.

2. Pourquoi une faible distance intra-cluster est-elle souhaitable ?

A. Elle indique que les points de données sont bien séparés des autres clusters.

B. Elle montre que les points de données dans un cluster sont très similaires entre eux.

C. Elle réduit le nombre de clusters nécessaires.

D. Elle augmente la taille des clusters.

3. Comment mesure-t-on la distance intra-cluster ?

A. Par la somme des carrés des distances entre les points de données et le centroïde du cluster.

B. Par la distance moyenne entre les points de données de différents clusters.

C. Par le nombre de points dans chaque cluster.

D. Par la densité des points dans le cluster.

4. Quelle est la signification de la distance inter-cluster ?

A. La similarité entre les points à l'intérieur d'un même cluster.

B. La distance entre les centroïdes de différents clusters.

C. La distance entre les points et le centroïde du même cluster.

D. La densité des points à l'intérieur des clusters.

5. Pourquoi une grande distance inter-cluster est-elle souhaitable ?

A. Elle indique que les points de données sont bien regroupés à l'intérieur des clusters.

B. Elle montre que les différents clusters sont bien séparés et distincts les uns des autres.

C. Elle réduit le nombre de points de bruit.

D. Elle augmente la cohésion intra-cluster.

6. Quel est l'objectif principal d'un bon clustering en termes de distances intra-cluster et inter-cluster ?

A. Minimiser la distance intra-cluster et maximiser la distance inter-cluster.

B. Maximiser la distance intra-cluster et minimiser la distance inter-cluster.

C. Minimiser les distances intra-cluster et inter-cluster.

D. Maximiser les distances intra-cluster et inter-cluster.

7. Quel métrique combine la cohésion intra-cluster et la séparation inter-cluster pour évaluer la qualité des clusters ?

A. Indice de Gini

B. Coefficient de silhouette

C. Entropie

D. Score de Dunn

8. Une grande distance intra-cluster indique :

A. Une bonne homogénéité au sein du cluster.

B. Une mauvaise homogénéité au sein du cluster.

C. Une bonne séparation entre les clusters.

D. Une grande densité de points dans le cluster.

9. La séparation inter-cluster peut être mesurée par :

A. La somme des carrés des distances entre les points de données et le centroïde du cluster.

B. La distance entre les centroïdes des clusters.

C. La distance moyenne entre les points de données du même cluster.

D. La densité des points à l'intérieur des clusters.

10. Quelle est une bonne pratique pour évaluer la qualité globale d'un clustering ?

A. Minimiser la cohésion intra-cluster tout en maximisant la séparation inter-cluster.

B. Maximiser la cohésion intra-cluster et la séparation inter-cluster.

C. Minimiser les distances intra-cluster et inter-cluster.

D. Maximiser les distances intra-cluster et inter-cluster.

11. Qu'est-ce qu'un faible score de cohésion intra-cluster suggère ?

A. Les points de données sont bien regroupés à l'intérieur du cluster.

B. Les points de données sont mal regroupés à l'intérieur du cluster.

C. Les clusters sont bien séparés.

D. Les clusters sont mal séparés.

12. Quelle méthode de visualisation peut aider à comprendre la structure des clusters ?

A. Diagramme de Venn

B. Diagramme de silhouettes

C. Boîte à moustaches

D. Histogramme

13. Pourquoi une bonne séparation inter-cluster est-elle importante ?

A. Elle réduit le nombre total de clusters.

B. Elle assure que les clusters sont bien distincts et non chevauchants.

C. Elle augmente la densité des points dans chaque cluster.

D. Elle réduit la variance totale des données.

14. Quel est l'objectif principal de l'analyse de la distance intra-cluster ?

A. Réduire la similarité entre les points de données.

B. Réduire la dissimilarité entre les points de données du même cluster.

C. Augmenter la séparation entre les clusters.

D. Augmenter le nombre total de clusters.

15. Que signifie un coefficient de silhouette proche de 1 ?

A. Les points sont mal assignés aux clusters.

B. Les clusters ne sont pas bien séparés.

C. Les points sont bien assignés et les clusters sont bien séparés.

D. Les clusters ont une grande densité.

16. Qu'est-ce qui pourrait indiquer une mauvaise évaluation de la qualité des clusters ?

A. Une faible distance intra-cluster et une faible distance inter-cluster.

B. Une faible distance intra-cluster et une grande distance inter-cluster.

C. Une grande distance intra-cluster et une grande distance inter-cluster.

D. Une grande distance intra-cluster et une faible distance inter-cluster.

17. Quel est le rôle des distances intra-cluster et inter-cluster dans le coefficient de silhouette ?

A. Mesurer uniquement la cohésion des clusters.

B. Mesurer uniquement la séparation des clusters.

C. Mesurer à la fois la cohésion des clusters et la séparation des clusters.

D. Ignorer la cohésion et se concentrer sur la séparation des clusters.

18. Comment le score de Dunn est-il utilisé dans l'évaluation des clusters ?

A. Il mesure la variance des points de données.

B. Il combine la cohésion intra-cluster et la séparation inter-cluster.

C. Il mesure la densité des points dans chaque cluster.

D. Il détermine le nombre optimal de clusters.

19. Une bonne homogénéité au sein des clusters est indiquée par :

A. Une faible distance intra-cluster.

B. Une grande distance intra-cluster.

C. Une faible distance inter-cluster.

D. Une grande distance inter-cluster.

20. Quel est l'avantage principal d'un bon équilibre entre la cohésion intra-cluster et la séparation inter-cluster ?

A. Il assure que les points de données sont bien assignés et que les clusters sont distincts.

B. Il augmente le nombre de clusters.

C. Il réduit le nombre de points de bruit.

D. Il améliore la densité des clusters.