ANN - Clustering #1

1. Quel algorithme de clustering nécessite de spécifier le nombre de clusters à l'avance ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Aucune des réponses ci-dessus

2. Quel algorithme de clustering est le plus efficace pour identifier des clusters de formes irrégulières ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Tous les trois

3. Quel algorithme de clustering génère un dendrogramme pour visualiser les relations entre les points de données ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. K-means et DBSCAN

4. Quel est l'avantage principal de K-means clustering ? (2 points)

- A. Peut gérer des clusters de formes irrégulières

- B. Génère un dendrogramme

- C. Simple à comprendre et à implémenter

- D. Insensible aux valeurs aberrantes

5. Quel algorithme de clustering est le plus sensible aux valeurs aberrantes et au bruit ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Aucun des trois

6. Quel algorithme de clustering n'a pas besoin de spécifier le nombre de clusters à l'avance ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Toutes les réponses

7. Quel est l'inconvénient principal du clustering hiérarchique ? (2 points)

- A. Doit spécifier le nombre de clusters à l'avance

- B. Sensible aux valeurs aberrantes

- C. Moins performant pour des clusters de densité variable

- D. Génère un dendrogramme

8. Quel est l'avantage principal de DBSCAN ? (2 points)

- A. Simple à comprendre et à implémenter

- B. Capable de gérer des valeurs aberrantes et du bruit

- C. Rapide pour des grands ensembles de données

- D. Génère un dendrogramme

9. Quel est le rôle de l'epsilon (e) dans l'algorithme DBSCAN ? (2 points)

- A. Détermine le nombre de clusters

- B. Détermine le rayon de voisinage autour d'un point

- C. Détermine le nombre minimum de points requis dans un rayon e

- D. Aucune des réponses ci-dessus

10. Quelle métrique est utilisée pour évaluer la qualité des clusters créés par un modèle de clustering ? (2 points)

- A. Score de silhouette

- B. Dendrogramme

- C. Centroides

- D. Matrice de confusion

11. Un score de silhouette proche de 1 indique que : (2 points)

- A. Les points de données sont mal assignés à un cluster incorrect

- B. Les points de données sont bien séparés des autres clusters

- C. Les points de données sont sur ou très près de la frontière de décision entre deux clusters voisins

- D. Les clusters sont globulaires

12. Quel algorithme de clustering fonctionne bien pour des clusters globulaires et bien séparés ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Aucune des réponses ci-dessus

13. Quel algorithme de clustering peut capturer des clusters de formes variées ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. B et C

14. Pour quels types de jeux de données DBSCAN est-il le plus performant ? (2 points)

- A. Jeux de données avec des clusters sphériques

- B. Jeux de données avec des valeurs aberrantes et des clusters de formes irrégulières

- C. Jeux de données sans valeurs aberrantes

- D. Jeux de données globulaires

15. Quel algorithme de clustering est plus lent et inefficace pour des grands ensembles de données ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Aucune des réponses ci-dessus

16. Quelle est l'importance de normaliser les données avant d'appliquer DBSCAN ? (2 points)

- A. Pour rendre les données plus lisibles

- B. Pour s'assurer que toutes les caractéristiques contribuent également au processus de clustering

- C. Pour augmenter la taille des données

- D. Aucune des réponses ci-dessus

17. Quel algorithme utilise des centroides pour définir les clusters ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Tous les trois

18. Pour quel algorithme le réglage des hyperparamètres est-il crucial ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Tous les trois

19. Quel algorithme est capable de détecter les points de bruit dans un jeu de données ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Tous les trois

20. Quelle méthode de linkage est utilisée dans le clustering hiérarchique pour minimiser la variance à l'intérieur des clusters ? (2 points)

- A. Single linkage

- B. Complete linkage

- C. Average linkage

- D. Ward's method

21. Quel paramètre du DBSCAN définit le nombre minimum de points nécessaires pour former un cluster ? (2 points)

- A. Epsilon (e)

- B. MinPts

- C. k

- D. Distance maximale

22. Quel algorithme de clustering utilise un critère de variance pour séparer les clusters ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Clustering par densité

23. Quelle est l'importance de la normalisation des données avant d'appliquer K-means ? (2 points)

- A. Pour augmenter la précision des centroides

- B. Pour s'assurer que chaque caractéristique contribue également à la distance euclidienne

- C. Pour réduire la taille des données

- D. Pour rendre les données plus lisibles

24. Quel algorithme de clustering est plus adapté pour détecter des outliers ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

25. Quelle est la complexité temporelle de l'algorithme K-means par itération ? (2 points)

- A. O(n)

- B. O(n log n)

- C. O(kn)

- D. O(n^2)

26. Dans le clustering hiérarchique, quel type de linkage peut provoquer des chaînes d'éléments ? (2 points)

- A. Single linkage

- B. Complete linkage

- C. Average linkage

- D. Ward's method

27. Quel algorithme de clustering est généralement utilisé en combinaison avec la réduction de dimensionnalité ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

28. Quel est l'inconvénient majeur de l'algorithme de clustering hiérarchique agglomératif ? (2 points)

- A. Sensible aux valeurs aberrantes

- B. Nécessite de spécifier le nombre de clusters à l'avance

- C. Ne génère pas de dendrogramme

- D. Ne peut pas gérer des clusters de formes irrégulières

29. Quel est le principal avantage de l'algorithme K-means++ par rapport à l'algorithme K-means classique ? (2 points)

- A. Génère des clusters plus sphériques

- B. Minimise la variance à l'intérieur des clusters

- C. Meilleure initialisation des centroides

- D. Plus rapide pour les grands ensembles de données

30. Quel algorithme de clustering est basé sur la notion de densité des points dans un espace multidimensionnel ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

31. Quelle est la méthode de linkage la plus appropriée pour le clustering hiérarchique lorsque les clusters sont de tailles inégales ? (2 points)

- A. Single linkage

- B. Complete linkage

- C. Average linkage

- D. Ward's method

32. Quelle est la principale limitation de l'alg

orithme K-means lorsqu'il s'agit de clusters de densité variable ? (2 points)

- A. Sensible aux valeurs aberrantes

- B. Doit spécifier le nombre de clusters à l'avance

- C. Ne peut pas détecter les outliers

- D. Suppose que les clusters ont la même densité

33. Quel algorithme de clustering est souvent utilisé avec les données spatiales pour détecter les clusters géographiques ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

34. Quel est l'objectif principal du score de silhouette dans le contexte du clustering ? (2 points)

- A. Déterminer le nombre optimal de clusters

- B. Évaluer la cohésion interne et la séparation des clusters

- C. Détecter les points de bruit dans les clusters

- D. Visualiser les relations entre les clusters

35. Quel algorithme de clustering est utilisé dans le code suivant ?

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters=3, random\_state=42)

kmeans.fit(data)

labels = kmeans.labels\_

- A. Clustering Hiérarchique

- B. DBSCAN

- C. K-means

- D. Mean Shift

36. Que fait ce code Python ?

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

scaled\_data = scaler.fit\_transform(data)

- A. Applique le clustering K-means

- B. Applique le clustering hiérarchique

- C. Normalise les données

- D. Identifie les outliers

37. Quel type de linkage est utilisé dans ce code de clustering hiérarchique ?

from scipy.cluster.hierarchy import linkage

Z = linkage(data, method='complete')

- A. Single linkage

- B. Complete linkage

- C. Average linkage

- D. Ward's method

38. Quelle fonction est utilisée pour générer un dendrogramme dans ce code ?

from scipy.cluster.hierarchy import dendrogram

dendrogram(Z)

- A. linkage

- B. fit

- C. predict

- D. dendrogram

39. Quel algorithme de clustering est appliqué dans ce code ?

from sklearn.cluster import AgglomerativeClustering

model = AgglomerativeClustering(n\_clusters=4, affinity='euclidean', linkage='ward')

model.fit(data)

labels = model.labels\_

- A. K-means

- B. DBSCAN

- C. Clustering Hiérarchique

- D. Mean Shift

40. Quel est le rôle de la méthode 'fit' dans ce code K-means ?

kmeans.fit(data)

- A. Prévoir les labels des clusters

- B. Normaliser les données

- C. Former le modèle de clustering

- D. Calculer les distances entre les points

41. Quelle méthode de linkage est utilisée dans ce code pour le clustering hiérarchique ?

from scipy.cluster.hierarchy import linkage

Z = linkage(data, method='average')

- A. Single linkage

- B. Complete linkage

- C. Average linkage

- D. Ward's method

42. Quelle bibliothèque est utilisée pour calculer le score de silhouette dans ce code ?

from sklearn.metrics import silhouette\_score

score = silhouette\_score(data, labels)

- A. numpy

- B. pandas

- C. scipy

- D. sklearn

43. Quel algorithme de clustering est utilisé dans ce code ?

from sklearn.cluster import DBSCAN

dbscan = DBSCAN(eps=0.5, min\_samples=5)

dbscan.fit(data)

labels = dbscan.labels\_

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

44. Quel est le rôle de l'objet 'StandardScaler' dans ce code ?

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

scaled\_data = scaler.fit\_transform(data)

- A. Effectuer le clustering K-means

- B. Réduire les dimensions des données

- C. Normaliser les données

- D. Identifier les outliers

45. Quel algorithme de clustering est utilisé dans ce code ?

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters=3, init='k-means++', random\_state=42)

kmeans.fit(data)

labels = kmeans.labels\_

- A. DBSCAN

- B. Clustering Hiérarchique

- C. K-means

- D. Mean Shift

46. Quel est le but de ce code ?

from scipy.cluster.hierarchy import dendrogram, linkage

Z = linkage(data, method='ward')

dendrogram(Z)

plt.show()

- A. Appliquer le clustering K-means

- B. Appliquer DBSCAN

- C. Créer un dendrogramme pour le clustering hiérarchique

- D. Calculer le score de silhouette

47. Quel algorithme de clustering est illustré dans ce code ?

from sklearn.cluster import AgglomerativeClustering

clustering = AgglomerativeClustering(n\_clusters=3, affinity='euclidean', linkage='ward')

clustering.fit(data)

labels = clustering.labels\_

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

48. Quel est le rôle de ce code ?

from sklearn.cluster import DBSCAN

dbscan = DBSCAN(eps=0.3, min\_samples=10)

dbscan.fit(data)

labels = dbscan.labels\_

- A. Appliquer le clustering K-means

- B. Appliquer le clustering hiérarchique

- C. Appliquer le clustering DBSCAN

- D. Appliquer le clustering Mean Shift

49. Quel est le rôle de ce code ?

from sklearn.metrics import silhouette\_score

score = silhouette\_score(data, labels)

print(f'Score de silhouette: {score}')

- A. Appliquer le clustering K-means

- B. Calculer le score de silhouette

- C. Créer un dendrogramme

- D. Normaliser les données

50. Quel algorithme de clustering est utilisé dans ce code ?

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters=4, random\_state=42)

kmeans.fit(data)

labels = kmeans.labels\_

- A. DBSCAN

- B. Clustering Hiérarchique

- C. K-means

- D. Mean Shift

51. Quel algorithme de clustering est illustré dans ce code ?

from sklearn.cluster import AgglomerativeClustering

model = AgglomerativeClustering(n\_clusters=2, affinity='euclidean', linkage='average')

model.fit(data)

labels = model.labels\_

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

52. Que fait ce code Python ?

from sklearn.cluster import DBSCAN

dbscan = DBSCAN(eps=0.7, min\_samples=15)

dbscan.fit(data)

labels = dbscan.labels\_

- A. Applique le clustering K-means

- B. Applique le clustering hiérarchique

- C. Applique le clustering DBSCAN

- D. Applique le clustering Mean Shift

53. Quel est le rôle de ce code ?

from sklearn.metrics import silhouette\_score

score = silhouette\_score(data, labels)

- A. Appliquer le clustering K-means

- B. Calculer le score de silhouette

- C. Créer un dendrogramme

- D. Normaliser les données

54. Quel algorithme de clustering est utilisé dans ce code ?

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters=5, random\_state=42)

kmeans.fit(data)

labels = kmeans.labels\_

- A. DBSCAN

- B. Clustering Hiérarchique

- C. K-means

- D. Mean Shift

ANN - Clustering #1 (Suite)

55. Quel algorithme de clustering est illustré dans ce code ?

from sklearn.cluster import AgglomerativeClustering

model = AgglomerativeClustering(n\_clusters=2, affinity='euclidean', linkage='average')

model.fit(data)

labels = model.labels\_

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

56. Que fait ce code Python ?

from sklearn.cluster import DBSCAN

dbscan = DBSCAN(eps=0.7, min\_samples=15)

dbscan.fit(data)

labels = dbscan.labels\_

- A. Applique le clustering K-means

- B. Applique le clustering hiérarchique

- C. Applique le clustering DBSCAN

- D. Applique le clustering Mean Shift

57. Quel est le rôle de ce code ?

from sklearn.metrics import silhouette\_score

score = silhouette\_score(data, labels)

- A. Appliquer le clustering K-means

- B. Calculer le score de silhouette

- C. Créer un dendrogramme

- D. Normaliser les données

58. Quel algorithme de clustering est utilisé dans ce code ?

from sklearn.cluster import KMeans

kmeans = KMeans(n\_clusters=5, random\_state=42)

kmeans.fit(data)

labels = kmeans.labels\_

- A. DBSCAN

- B. Clustering Hiérarchique

- C. K-means

- D. Mean Shift

Ce contenu n'a pas été créé ni n’est approuvé par Microsoft. Les données que vous soumettez sont envoyées au propriétaire du formulaire. Microsoft Forms

59. Quelle méthode de clustering est utilisée principalement pour la visualisation des clusters ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

60. Quel algorithme de clustering est basé sur la notion de densité des points dans un espace multidimensionnel ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

61. Quel algorithme de clustering peut gérer des ensembles de données de haute dimension sans normalisation ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

62. Quel type de clustering utilise l'analyse de densité locale pour former des clusters ? (2 points)

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift

63. Quel algorithme est implémenté dans ce code pour détecter les clusters de densité ?

from sklearn.cluster import DBSCAN

dbscan = DBSCAN(eps=0.5, min\_samples=5)

dbscan.fit(data)

labels = dbscan.labels\_

- A. K-means

- B. Clustering Hiérarchique

- C. DBSCAN

- D. Mean Shift