**Quiz 2 sur l'Optimisation de l'Apprentissage Non Supervisé - DBSCAN**

Partie 1 - Théorie de DBSCAN

1. Quelle est la principale caractéristique de l'algorithme DBSCAN ?

A. Il nécessite de spécifier le nombre de clusters à l'avance.

B. Il peut trouver des clusters de formes arbitraires et est robuste aux bruits.

C. Il est sensible aux valeurs aberrantes.

D. Il n'est pas adapté pour les ensembles de données de grande taille.

2. Dans DBSCAN, que représente le paramètre Epsilon (ε) ?

A. La distance maximale entre deux points pour qu'ils soient considérés comme voisins.

B. Le nombre minimum de points requis pour former un cluster.

C. La distance minimale entre deux points pour qu'ils soient considérés comme voisins.

D. Le nombre maximum de points requis pour former un cluster.

3. Quel est le rôle du paramètre minPts dans DBSCAN ?

A. Déterminer la taille maximale d'un cluster.

B. Déterminer la distance maximale entre deux points.

C. Déterminer le nombre minimum de points requis pour former une région dense.

D. Déterminer la distance minimale entre deux points.

4. Que se passe-t-il si un point a moins de minPts voisins dans un rayon de ε ?

A. Il est considéré comme un point de cœur.

B. Il est considéré comme un point de bordure.

C. Il est considéré comme du bruit.

D. Il est automatiquement assigné au cluster le plus proche.

5. Pourquoi DBSCAN est-il considéré comme robuste face aux valeurs aberrantes ?

A. Parce qu'il nécessite de spécifier le nombre de clusters.

B. Parce qu'il ignore les points qui ne sont pas suffisamment denses pour former un cluster.

C. Parce qu'il est basé sur la distance euclidienne.

D. Parce qu'il ne peut pas trouver des clusters de formes arbitraires.

6. Quelle est la première étape de l'algorithme DBSCAN ?

A. Identifier les points de cœur.

B. Former des clusters.

C. Définir les paramètres ε et minPts.

D. Identifier le bruit.

7. Comment DBSCAN traite-t-il les points qui ne sont ni des points de cœur ni accessibles depuis un point de cœur ?

A. Ils sont marqués comme des points de bordure.

B. Ils sont marqués comme du bruit.

C. Ils sont fusionnés avec le cluster le plus proche.

D. Ils sont ignorés.

8. Quel est le rôle des points de cœur dans DBSCAN ?

A. Ils servent de centres de clusters.

B. Ils sont utilisés pour marquer les points de bruit.

C. Ils sont connectés pour former des clusters.

D. Ils déterminent la distance maximale entre deux points.

9. Que signifie "MinPts" dans le contexte de DBSCAN ?

A. Le nombre de points dans le plus petit cluster.

B. Le nombre minimum de points nécessaires pour former un cluster dense.

C. La distance minimale entre deux points dans un cluster.

D. La densité maximale d'un cluster.

10. Comment DBSCAN étend-il un cluster à partir d'un point de cœur ?

A. En ajoutant seulement les points voisins directs.

B. En ajoutant tous les voisins accessibles depuis les points de cœur.

C. En fusionnant les clusters voisins.

D. En utilisant une méthode hiérarchique pour relier les points.

Partie 2 - Paramètres de DBSCAN

11. Quelle est l'importance de bien choisir la valeur de ε ?

A. Elle détermine la taille maximale des clusters.

B. Elle influence la densité nécessaire pour former un cluster.

C. Elle détermine le nombre de points marqués comme du bruit.

D. Elle n'a pas d'impact significatif sur les résultats de DBSCAN.

12. Que se passe-t-il si ε est trop petit ?

A. Il y aura trop de points de bruit et peu de clusters.

B. Tous les points seront regroupés en un seul cluster.

C. Les clusters seront trop grands.

D. Les clusters seront mal définis et non compacts.

13. Que se passe-t-il si ε est trop grand ?

A. Il n'y aura aucun point de bruit.

B. Tous les points seront considérés comme des points de cœur.

C. Les clusters distincts seront fusionnés en un seul.

D. Les clusters seront trop petits.

14. Pourquoi est-il conseillé de fixer MinPts à au moins la dimension du jeu de données plus un ?

A. Pour garantir que les clusters sont suffisamment denses.

B. Pour éviter que trop de points soient marqués comme du bruit.

C. Pour assurer la robustesse face aux valeurs aberrantes.

D. Pour éviter que les clusters soient trop grands.

15. Quelle méthode peut être utilisée pour déterminer la valeur optimale de ε ?

A. La méthode de la coudée.

B. Le diagramme des distances k.

C. L'analyse de la variance.

D. La validation croisée.

16. Quel est le but du diagramme des distances k dans DBSCAN ?

A. Identifier la distance minimale entre les points.

B. Déterminer le nombre optimal de clusters.

C. Trouver le point de courbure pour choisir ε.

D. Calculer la densité maximale des clusters.

17. Quelle est une règle empirique pour fixer MinPts ?

A. MinPts doit être égal à la dimension des données moins un.

B. MinPts doit être fixé à 10 pour tous les jeux de données.

C. MinPts doit être au moins la dimension des données plus un.

D. MinPts doit être la moitié de la taille totale des données.

18. Comment DBSCAN traite-t-il les valeurs aberrantes ?

A. Les valeurs aberrantes sont incluses dans les clusters.

B. Les valeurs aberrantes sont marquées comme du bruit.

C. Les valeurs aberrantes sont fusionnées avec les points de bordure.

D. Les valeurs aberrantes sont ignorées.

19. Pourquoi DBSCAN est-il adapté aux clusters de formes variées ?

A. Parce qu'il utilise des centres de clusters fixes.

B. Parce qu'il ne nécessite pas de spécifier le nombre de clusters.

C. Parce qu'il ignore les points de bruit.

D. Parce qu'il utilise des distances minimales pour former des clusters.

20. Quelle est l'une des limitations de DBSCAN ?

A. Il nécessite de spécifier le nombre de clusters.

B. Il ne peut pas gérer les valeurs aberrantes.

C. Il est sensible aux paramètres ε et minPts.

D. Il ne peut pas trouver des clusters de formes arbitraires.

Partie 3 - Méthodes pour Optimiser les Paramètres

21. Quelle méthode est couramment utilisée pour visualiser la valeur optimale de ε ?

A. Diagramme de Pareto.

B. Courbe ROC.

C. Diagramme des distances k.

D. Courbe de Lorenz.

22. Que représente le "coude" dans le diagramme des distances k ?

A. La distance maximale entre les points.

B. Le nombre optimal de clusters.

C. La valeur optimale de ε.

D. Le nombre minimal de points nécessaires pour former un cluster.

23. Quelle est l'étape suivante après avoir trouvé le "coude" dans le diagramme des distances k ?

A. Ajuster MinPts en conséquence.

B. Appliquer DBSCAN avec la valeur optimale de ε.

C. Déterminer le nombre de clusters.

D. Marquer les points de bruit.

24. Pourquoi est-il important de normaliser les données avant d'appliquer DBSCAN ?

A. Pour réduire le nombre de clusters.

B. Pour améliorer la précision du clustering.

C. Pour rendre les distances comparables entre les dimensions.

D. Pour ignorer les points de bruit.

25. Quelle est une approche alternative pour optimiser ε et MinPts ?

A. Utiliser la méthode de la validation croisée.

B. Utiliser la régression linéaire.

C. Utiliser la réduction de dimensionnalité.

D. Utiliser l'analyse en composantes principales (PCA).

26. Que se passe-t-il si les données ne sont pas normalisées avant d'appliquer DBSCAN ?

A. Les distances entre les points seront faussées.

B. Les clusters seront plus compacts.

C. Le nombre de points de bruit sera réduit.

D. Les clusters seront mieux définis.

27. Quelle est une des difficultés de l'optimisation des paramètres de DBSCAN ?

A. Trouver le nombre de clusters à l'avance.

B. Déterminer les points de bordure.

C. Identifier les points de bruit.

D. Choisir les valeurs optimales de ε et MinPts.

28. Pourquoi utiliser des heuristiques pour choisir MinPts ?

A. Pour simplifier le processus de clustering.

B. Pour garantir des clusters de taille égale.

C. Pour éviter les valeurs aberrantes.

D. Pour assurer une densité minimale des clusters.

29. Quelle est l'influence de la dimension des données sur le choix de MinPts ?

A. Plus la dimension est élevée, plus MinPts doit être faible.

B. Plus la dimension est élevée, plus MinPts doit être élevé.

C. La dimension des données n'a pas d'influence sur MinPts.

D. La dimension des données influe seulement sur ε.

30. Quelle est une méthode pour évaluer la qualité des clusters formés par DBSCAN ?

A. Utiliser le coefficient de silhouette.

B. Utiliser la régression logistique.

C. Utiliser l'analyse discriminante linéaire.

D. Utiliser l'analyse factorielle.

Partie 4 - Exemples Pratiques

31. Quel est l'impact d'un ε trop grand dans DBSCAN ?

A. Il n'y aura aucun point de bruit.

B. Tous les points seront marqués comme du bruit.

C. Les clusters distincts seront fusionnés en un seul.

D. Les clusters seront trop compacts.

32. Quel est l'impact d'un ε trop petit dans DBSCAN ?

A. Tous les points seront regroupés en un seul cluster.

B. Les points seront marqués comme des points de cœur.

C. Il y aura trop de points de bruit et peu de clusters.

D. Les clusters seront trop grands.

33. Comment vérifier si les valeurs de ε et MinPts choisies sont optimales ?

A. En utilisant une courbe d'inertie.

B. En calculant l'indice de Davies-Bouldin.

C. En utilisant l'analyse en composantes principales.

D. En calculant le coefficient de silhouette.

34. Pourquoi est-il important de visualiser les résultats du clustering avec DBSCAN ?

A. Pour identifier le nombre exact de clusters.

B. Pour ajuster les paramètres en fonction des résultats.

C. Pour déterminer les points de cœur.

D. Pour ignorer les points de bruit.

35. Quelle est l'étape finale de l'application de DBSCAN après avoir optimisé les paramètres ?

A. Réduire la dimensionnalité des données.

B. Marquer les points de bruit.

C. Évaluer la qualité des clusters formés.

D. Normaliser les données.

36. Quelle est une approche courante pour tester différentes valeurs de ε et MinPts ?

A. Utiliser une grille de recherche.

B. Utiliser la méthode de la coudée.

C. Utiliser la régression logistique.

D. Utiliser l'analyse factorielle.

37. Quel est l'avantage principal de DBSCAN par rapport à K-means ?

A. Il nécessite de spécifier le nombre de clusters.

B. Il peut trouver des clusters de formes variées sans connaître le nombre de clusters.

C. Il est sensible aux valeurs aberrantes.

D. Il utilise des centres de clusters fixes.

38. Comment DBSCAN traite-t-il les points de bruit ?

A. Ils sont fusionnés avec les clusters voisins.

B. Ils sont marqués comme des points de cœur.

C. Ils sont ignorés et exclus des clusters.

D. Ils sont ajoutés aux clusters existants.

39. Pourquoi DBSCAN est-il bien adapté pour les données GPS ?

A. Parce qu'il ignore les points de bruit.

B. Parce qu'il peut détecter des clusters de formes variées sans connaître le nombre de clusters.

C. Parce qu'il utilise des centres de clusters fixes.

D. Parce qu'il nécessite de spécifier le nombre de clusters.

40. Quelle est une limitation de l'utilisation de DBSCAN pour des données de haute dimension ?

A. Il ne peut pas trouver des clusters de formes arbitraires.

B. Il est sensible aux valeurs aberrantes.

C. La distance devient moins significative en haute dimension.

D. Il nécessite de spécifier le nombre de clusters.

Partie 5 - Applications et Cas d'Usage

41. Comment DBSCAN peut-il être utilisé dans l'analyse des données de transactions financières ?

A. Pour regrouper les transactions par type.

B. Pour détecter les transactions frauduleuses en tant que points de bruit.

C. Pour prédire les futures transactions.

D. Pour calculer la moyenne des transactions par client.

42. Quelle est une application de DBSCAN dans le domaine de la biologie ?

A. Pour séquencer des gènes.

B. Pour analyser des images microscopiques.

C. Pour regrouper des espèces similaires sur la base de caractéristiques génétiques.

D. Pour modéliser la croissance des cellules.

43. Pourquoi DBSCAN est-il utile pour l'analyse des réseaux sociaux ?

A. Pour prédire les interactions futures entre les utilisateurs.

B. Pour identifier des communautés d'utilisateurs similaires.

C. Pour calculer le score de popularité des utilisateurs.

D. Pour suivre les tendances des interactions en temps réel.

44. Comment DBSCAN peut-il aider à l'analyse des données météorologiques ?

A. Pour prédire les futures conditions météorologiques.

B. Pour regrouper des régions avec des conditions météorologiques similaires.

C. Pour calculer la moyenne des températures.

D. Pour analyser les tendances des précipitations.

45. Quelle est une utilisation de DBSCAN dans l'industrie de la vente au détail ?

A. Pour prévoir les ventes futures.

B. Pour regrouper les clients en fonction de leurs comportements d'achat.

C. Pour analyser les tendances des ventes.

D. Pour calculer le revenu moyen par client.

46. Pourquoi DBSCAN est-il adapté à l'analyse des données d'image ?

A. Parce qu'il nécessite de spécifier le nombre de clusters.

B. Parce qu'il peut détecter des objets de formes variées dans les images.

C. Parce qu'il ignore les points de bruit.

D. Parce qu'il utilise des centres de clusters fixes.

47. Comment DBSCAN peut-il être utilisé dans le domaine de la santé ?

A. Pour prédire les maladies futures.

B. Pour regrouper des patients ayant des conditions médicales similaires.

C. Pour analyser les tendances de consommation de médicaments.

D. Pour calculer le coût moyen des soins.

48. Quelle est une application de DBSCAN dans l'industrie de l'énergie ?

A. Pour prédire la consommation future d'énergie.

B. Pour regrouper des zones avec des niveaux de consommation d'énergie similaires.

C. Pour analyser les tendances de production d'énergie.

D. Pour calculer la consommation moyenne d'énergie par foyer.

49. Pourquoi DBSCAN est-il utile pour l'analyse des données de capteurs ?

A. Parce qu'il ignore les points de bruit.

B. Parce qu'il peut détecter des motifs de formes variées dans les données des capteurs.

C. Parce qu'il utilise des centres de clusters fixes.

D. Parce qu'il nécessite de spécifier le nombre de clusters.

50. Comment DBSCAN peut-il aider à l'analyse des données de transport ?

A. Pour prédire les temps de trajet futurs.

B. Pour regrouper des trajets similaires en fonction de leur durée et distance.

C. Pour analyser les tendances des accidents de la route.

D. Pour calculer la vitesse moyenne des véhicules.

Partie 6 - Avantages et Inconvénients de DBSCAN

51. Quel est l'avantage principal de DBSCAN par rapport à K-means ?

A. Il nécessite de spécifier le nombre de clusters.

B. Il peut trouver des clusters de formes variées sans connaître le nombre de clusters.

C. Il est sensible aux valeurs aberrantes.

D. Il utilise des centres de clusters fixes.

52. Pourquoi DBSCAN est-il robuste aux valeurs aberrantes ?

A. Parce qu'il les inclut dans les clusters.

B. Parce qu'il les marque comme du bruit et les ignore.

C. Parce qu'il les fusionne avec les points de bordure.

D. Parce qu'il utilise une méthode hiérarchique pour les traiter.

53. Quel est un inconvénient de DBSCAN ?

A. Il nécessite de spécifier le nombre de clusters.

B. Il est sensible aux choix des paramètres ε et MinPts.

C. Il ne peut pas gérer les valeurs aberrantes.

D. Il utilise des centres de clusters fixes.

54. Pourquoi le choix des paramètres ε et MinPts est-il crucial pour DBSCAN ?

A. Parce qu'ils déterminent la taille des clusters.

B. Parce qu'ils influencent la formation et la qualité des clusters.

C. Parce qu'ils définissent la distance entre les points.

D. Parce qu'ils n'ont pas d'impact sur les résultats.

55. Comment DBSCAN traite-t-il les points de bordure ?

A. Ils sont ignorés et marqués comme du bruit.

B. Ils sont fusionnés avec les clusters voisins.

C. Ils sont inclus dans le cluster du point de cœur le plus proche.

D. Ils forment de nouveaux clusters.

56. Pourquoi DBSCAN est-il adapté pour les grands ensembles de données ?

A. Parce qu'il nécessite de spécifier le nombre de clusters.

B. Parce qu'il peut trouver des clusters de formes variées sans connaître le nombre de clusters

C. Parce qu'il ignore les points de bruit.

D. Parce qu'il utilise des centres de clusters fixes.

57. Quel est l'avantage de DBSCAN pour les données spatiales ?

A. Il utilise des centres de clusters fixes.

B. Il peut trouver des clusters de formes variées sans connaître le nombre de clusters.

C. Il est sensible aux valeurs aberrantes.

D. Il nécessite de spécifier le nombre de clusters.

58. Pourquoi DBSCAN peut-il être inefficace pour les données de haute dimension ?

A. Il ne peut pas gérer les valeurs aberrantes.

B. La distance devient moins significative en haute dimension.

C. Il nécessite de spécifier le nombre de clusters.

D. Il utilise des centres de clusters fixes.

59. Comment DBSCAN peut-il être combiné avec d'autres techniques pour améliorer le clustering ?

A. En utilisant la régression linéaire.

B. En utilisant la réduction de dimensionnalité comme PCA avant l'application de DBSCAN.

C. En utilisant la méthode de la coudée.

D. En utilisant l'analyse factorielle.

60. Quelle est une limitation de DBSCAN lors de l'analyse de données avec des densités variables ?

A. Il peut mal identifier des clusters de densités très différentes.

B. Il ne peut pas trouver des clusters de formes arbitraires.

C. Il nécessite de spécifier le nombre de clusters.

D. Il ne peut pas gérer les valeurs aberrantes.