

1. Qual é a principal característica que define um cluster no DBScan?

- A principal característica é que um cluster é definido como um grupo de pontos que estão densamente conectados, ou seja, cada ponto dentro do cluster tem um número mínimo de vizinhos (minPts) dentro de um raio ϵ (epsilon).

2. O que significa um ponto ser classificado como "ruído" no DBScan?

- Um ponto é classificado como "ruído" se ele não estiver suficientemente próximo de nenhum outro ponto para formar parte de um cluster, ou seja, não tem o número mínimo de vizinhos dentro do raio ϵ .

3. Qual é a função principal do parâmetro ϵ (epsilon) no DBScan?

- O parâmetro ϵ define o raio de busca ao redor de cada ponto para encontrar vizinhos. É a distância máxima para considerar um ponto como vizinho de outro.

4. Qual é a função principal do parâmetro minPts no DBScan?

- O parâmetro minPts define o número mínimo de pontos necessários dentro do raio ϵ para que um ponto seja considerado como um "ponto central" e para que uma região seja considerada um cluster.

5. Em DBScan, quais tipos de pontos são classificados como "pontos de borda"?

- Pontos de borda são pontos que estão dentro do raio ϵ de um ponto central, mas não têm pontos suficientes em seu próprio raio ϵ para serem considerados pontos centrais.

6. Cite uma vantagem do DBScan em comparação com algoritmos de clusterização baseados em centroides, como K-means?

- Uma vantagem do DBScan é que ele pode identificar clusters de formas arbitrárias e lidar com ruído e outliers, enquanto K-means tende a encontrar clusters esféricos e é sensível a outliers.

7. DBScan pode ser aplicado a dados com ruído e outliers?

- Sim, DBScan é projetado para lidar com ruído e outliers, classificando-os como pontos de ruído, enquanto identifica clusters com base na densidade.

8. Qual é a complexidade computacional do DBScan em relação ao número de pontos N e à dimensão D dos dados?

- A complexidade computacional do DBScan é $O(N^2)$ no pior caso, mas pode ser melhorada para $O(N \log N)$ com o uso de estruturas de dados como árvores de busca espacial.

9. Explique o conceito de "ponto central" no DBScan e como ele contribui para a formação de clusters.

- Um "ponto central" é um ponto que tem pelo menos minPts vizinhos dentro do raio ϵ . Ele serve como o núcleo de um cluster, a partir do qual os pontos de borda e outros pontos densamente conectados são incluídos no cluster.

10. Descreva um cenário em que o DBScan seria uma escolha apropriada de algoritmo de clusterização. Inclua detalhes sobre a natureza dos dados e os parâmetros que devem ser ajustados.

- O DBScan seria apropriado em um cenário com dados que possuem clusters de formas variadas e incluem ruído, como em dados geoespaciais de áreas urbanas. Ajuste o parâmetro ϵ para refletir a densidade dos clusters e minPts para garantir a inclusão mínima de pontos para formar um cluster.

11. Como o DBScan lida com clusters de diferentes formas e tamanhos? Forneça exemplos e explique a eficácia do DBScan nesses casos.

- DBScan lida bem com clusters de diferentes formas e tamanhos, pois se baseia na densidade e não na forma geométrica dos clusters. Por exemplo, ele pode identificar clusters em forma de anel ou em formas irregulares. Sua eficácia decorre do fato de que não assume uma forma específica para os clusters.

12. Considere um conjunto de dados onde os pontos estão distribuídos em círculos concêntricos, com diferentes densidades. Discuta como o DBScan pode ser configurado para identificar corretamente esses clusters e quais desafios podem surgir.

- Para identificar círculos concêntricos com diferentes densidades, configure ϵ de forma que cubra a densidade esperada do círculo interno e ajuste minPts para distinguir entre a densidade dos círculos e o ruído. Desafios podem incluir a definição de ϵ apropriado para clusters de densidades variadas, que pode exigir ajuste fino e testes com diferentes valores para garantir a detecção adequada dos clusters concêntricos.