

Respuestas a las preguntas teóricas

a)

R: La situación descrita, para K-Means se podría dar si los datos estan muy cercanos entre sí y los valores que toma la función de distancia escogida entrega valores similares para todos los puntos, en otras palabras, se puede dar si hay solo una zona de alta densidad de los datos.

Con respecto a DBSCAN, este algoritmo está basado en densidad, y al utilizar single-link (menor distancia entre clusters) probablemente se dé el caso de que se detectará un solo cluster que contenga a la mayoría de los demás puntos, mientras que los puntos más lejanos podrían ser detectados como ruido.

b)

R: En el caso de single-link si es posible, la razón es que al utilizar hierarchival agglomerative clustering e iniciar con un cluster para cada objeto, luego de algunas aglomeraciones un objeto puede llegar a pertenecer a un cluster donde los objetos mas lejanos a él (en ese cluster) son en efecto más lejanos que otro objeto en otro cluster, solamente porque en cada aglomeración los demás objetos del cluster del punto en cuestión estaban más cerca de un objeto de otro cluster.

Para el caso de complete link, esto no es posible ya que el cluster/objeto habría decidido aglomerarse con el objeto más cercano.