Interactive Interview Questions

Clustering & MF

Tags:

#easy

#medium

#hard

#MLS

#MLE

#DS

(Q1) Explain / Draw failure cases of K-Means

#easy

(Q2) Psuedo code for computing Dunn-Index between m-clusters

#medium #MLE #DS

Inputs:

outputs:

Algorithm:

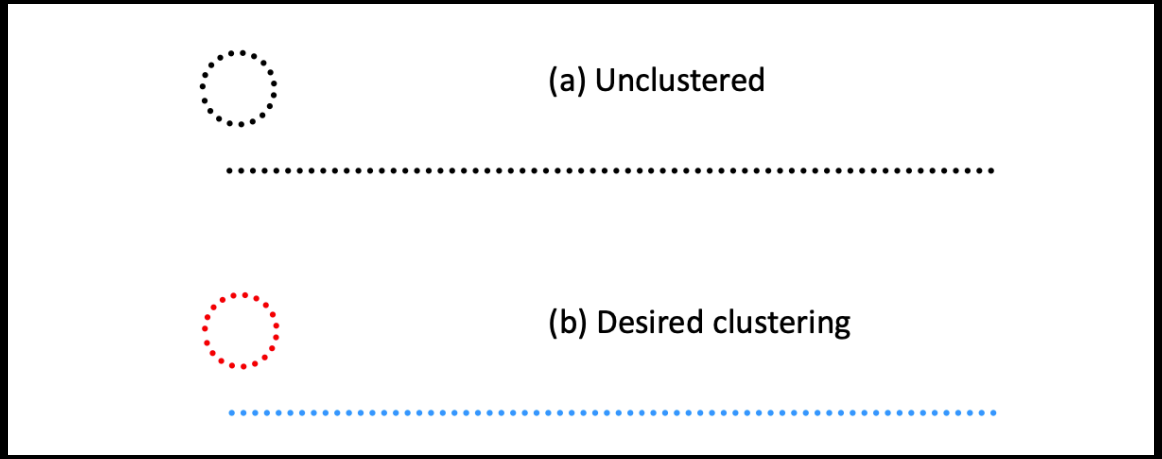
(Q3) (a) Time complexity to cluster using PAM algo for k-medoids. Let n = #points & k = #clusters. & ignore d = dimensionality of each point. f(n, k)

#easy #MLE #DS

(b) How would you speed it up?

#medium #DS

(Q4) Which of the following methods will cluster the data in panel (a) of the figure below into the two clusters (red circle and blue horizontal line) shown in panel (b)? Every dot in the circle and the line is a data point. In all the options that involve hierarchical clustering, the algorithm is run until we obtain two clusters. **[#Easy]**



A: Hierarchical agglomerative clustering with Euclidean distance and complete [=MAX] linkage
B: Hierarchical agglomerative clustering with Euclidean distance and single [=MIN] linkage
C: Hierarchical agglomerative clustering with Euclidean distance and centroid [=distance between centroids] linkage
D: *k*-means clustering with $k = 2$

Source: https://people.eecs.berkeley.edu/~jrs/189/

(Q5) Solve k-means clustering using SGD

#hard #MLS

(Q6) Failure cases of DB-scan. Justify.

#easy #DS

(Q1) 1 Billion points of 10-dim each. Cluster them on a single box efficiently & fast

#medium

#MLS

(Q8) 1 million docs & 50 categories

Manually label them with minimal human effort

#DS #medium

(Q9) How to optimize range-Queries in DB-scan

#medium #MLE

(Q10) KMeans ++ time complexity #easy #MLE