Hierarchical Clustering Quiz Questions

Zeqing Jin
zjin2017@berkeley.edu

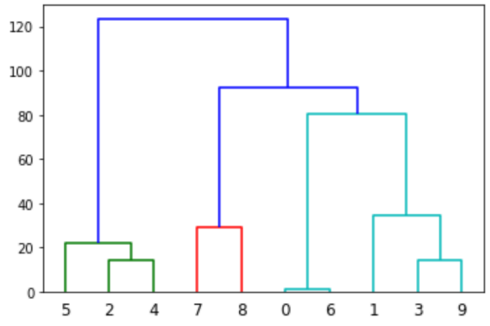
Xianlin Shao
shayd@berkeley.edu

Yifei Zhang
yifei_zhang@berkeley.edu

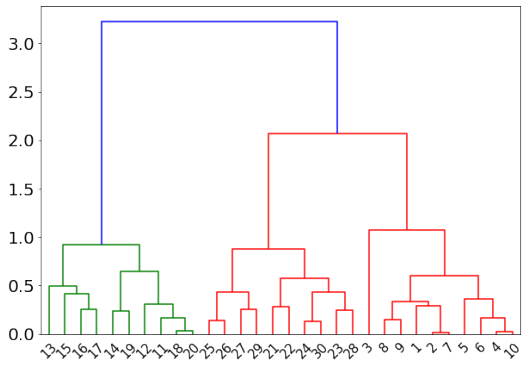
Zilan Zhang
shilan@berkeley.edu

Quiz Questions

1. Describe a scenario in which you can apply hierarchical clustering. State your objective, data, definition of affinity, and so on.
2. Which of the following statements are correct with respect to Hierarchical Clustering?
   A. Hierarchical clustering always requires to specify the number of clusters.
   B. Hierarchical clustering has a global optimization objective.
   C. Divisive clustering may require higher time complexity compared to aggomerative clustering
3. For naive agglomerative clustering, calculate the time complexity.
4. Answer the following questions based on the given Dendrogram.



   i. Which two clusters are merged in the first iteration?
   ii. If we break the cluster at dissimilarity level 60, how many clusters do we have ?
   iii. What is the approximate dissimilarity between cluster containing 0, 6 and cluster containing 1, 3, 9?
5. A dendrogram of hierarchical clustering using average linkage is given below. Roughly determine a reasonable number of cluster k.



6. The dissimilarity matrix of 5 points {1,2,3,4,5} is given below. Obviously we will first merge {3,4} because of minimum distance. Determine the two clusters merging in the next iteration using:

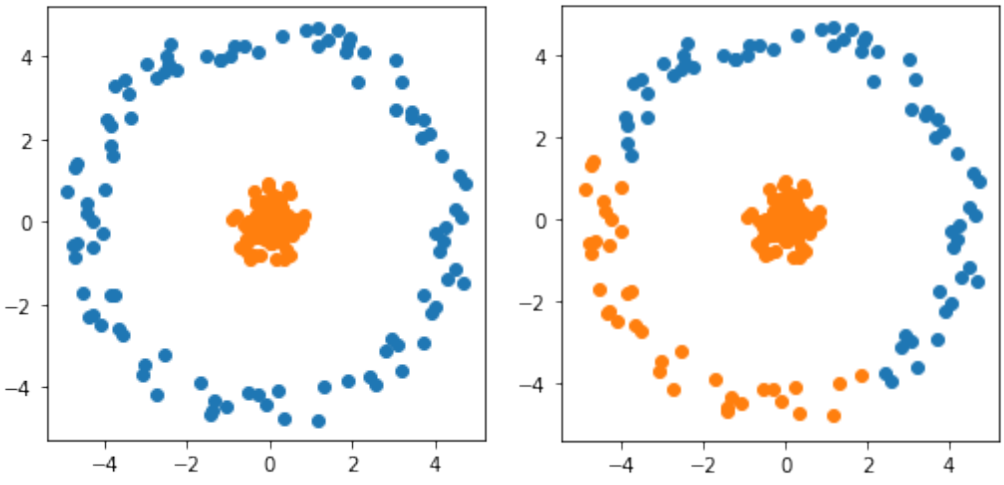
$$\begin{bmatrix} 0 & 12 & 40 & 35 & 24 \\ 12 & 0 & 21 & 66 & 27 \\ 40 & 21 & 0 & 8 & 18 \\ 35 & 66 & 8 & 0 & 10 \\ 24 & 27 & 18 & 10 & 0 \end{bmatrix}$$

   i. Single linkage.

   ii. Complete linkage.

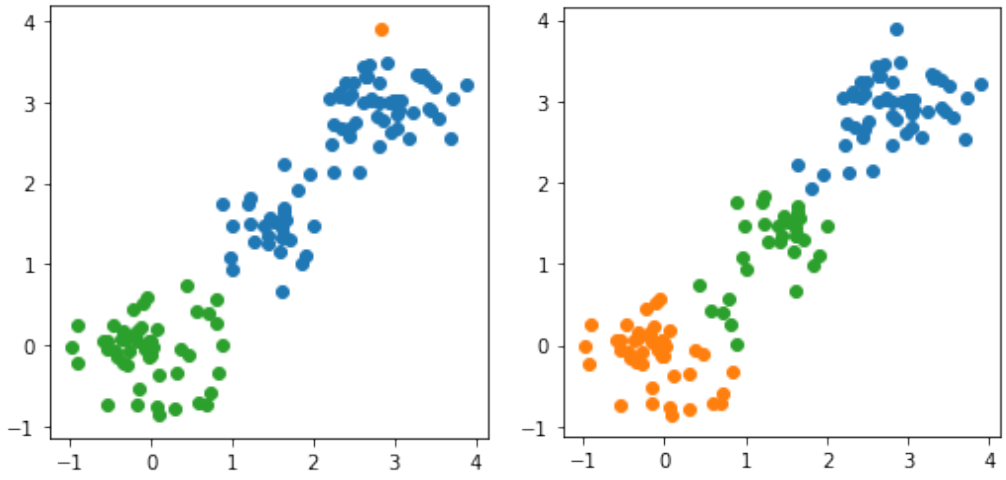
   iii. average linkage.

7. Determine which of the following clustering results ($k = 2$) is using single linkage and which one is using complete linkage.



(a) (b)

8. Determine which of the following clustering results ($k = 3$) is using single linkage and which one is using complete linkage.



(a) (b)

9. Which of the following statements are correct with respect to Divisive Hierarchical Clustering?
   A. Divisive Hierarchical clustering splits n objects from 1 cluster into n clusters.
   B. When deciding which cluster to split, we choose the one with most objects.
   C. When splitting one cluster, the splinter cluster always starts from one object.
10. (Optional) Which of the following statements are correct with respect to CURE?
   A. CURE has a good application on data with outliers.
   B. The shrinking factor moves all the points in a cluster to the centroid.
   C. CURE effectively reduce the time complexity compared to agglomerative clustering and divisive clustering.