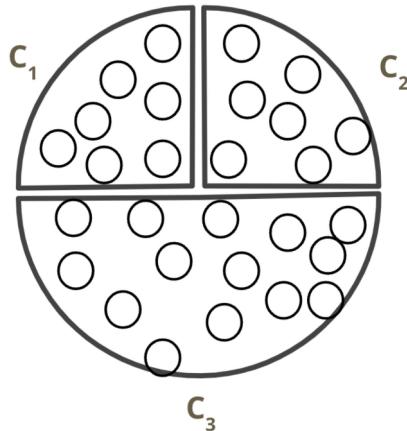


# CS 506 Lab 4 Questions

February 2026

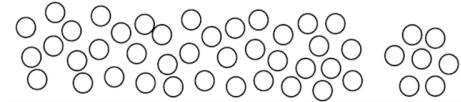
## Problems

1. Is the following a possible output of K-means++ clustering?

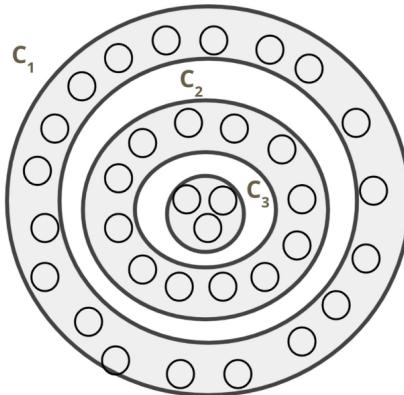


- (a) Yes  
(b) No
2. To find the optimal number of clusters in K-means, we run K-means multiple times for different values of  $k$  then we pick the one with the lowest cost
  - (a) True  
(b) False
3. What does it mean if the silhouette score for a data point is close to 1?
  - (a) The cluster is likely tight and far from others  
(b) The point lies on the boundary of clusters  
(c) The point does not belong in its cluster

4. Let  $k = 2$ . Is K-means++ with a Euclidean distance-based cost function ideal for the following data?

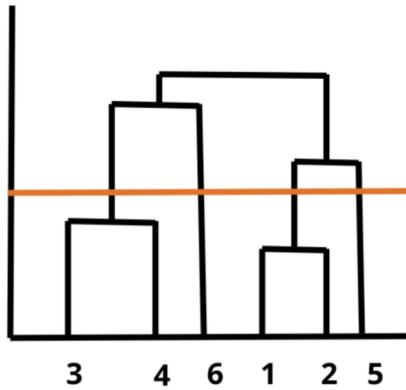


- (a) Yes, there are clearly two clusters here  
(b) No, we should use manhattan distance instead to account for the elongated cluster  
(c) No, k-means performs poorly on this data set because it assumes roughly spherical and equal sized clusters
5. You transform a 2D dataset of concentric circles onto a 1D space using  $z = (x - \bar{x})^2 + (y - \bar{y})^2$ . Would k-means be able to identify 3 concentric clusters in this new 1 dimensional space?



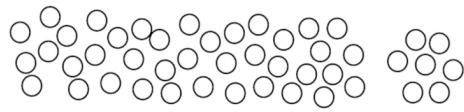
- (a) Yes  
(b) No

6. How many clusters would be created by cutting the dendrogram as below?



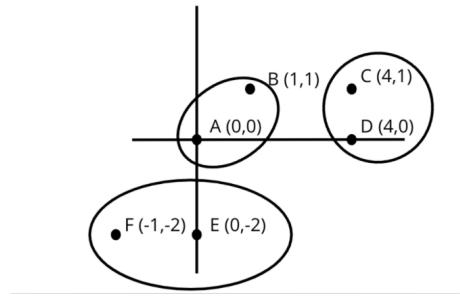
- (a) 7
- (b) 6
- (c) 5
- (d) 4

7. Which link function(s) would make Hierarchical clustering well suited for the following dataset?

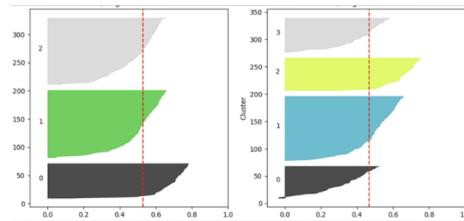


- (a) Single-link distance
- (b) Complete-link distance
- (c) Ward's distance
- (d) All of the above

8. Which clusters get merged if using Euclidean and complete-link distance?



- (a) AB and CD
  - (b) CD and EF
  - (c) AB and EF
9. Using Euclidean and complete-link distance for  $A = (1, 2)$ ,  $B = (3, 5)$ ,  $C = (4, 1)$ ,  $D = (7, 3)$ ,  $E = (6, 6)$  what is the merging order?
- (a)  $AC \rightarrow DE \rightarrow BDE \rightarrow ABCDE$
  - (b)  $AC \rightarrow DE \rightarrow ABC \rightarrow ABCDE$
  - (c)  $AC \rightarrow ABC \rightarrow DE \rightarrow ABCDE$
10. How would you compare the clustering represented by the left plot versus the right plot? (select all that apply)



- (a) the left plot represents 3 clusters, the right plot represents 4 clusters
- (b) the right plot has more data than the left plot
- (c) the gray cluster (top one) on the left plot has more points than the gray cluster (top one) on the right
- (d) the left plot contains at least one point with a negative silhouette score

11. We TYPICALLY prefer clusterings with higher average silhouette scores and uniformly distributed silhouette scores across data points
  - (a) True
  - (b) False
12. What is the average silhouette score for a clustering with  $k = N$ ?
  - (a) 0
  - (b) 1
  - (c)  $\infty$
  - (d)  $\sqrt{k}$

## Solutions

1. (b) No
2. (b) False
3. (a) The cluster is likely tight and far from others
4. (c) No, k-means performs poorly on this data set because it assumes roughly spherical and equal sized clusters
5. (a) Yes
6. (d) 4
7. (a) Single-link distance
8. (c) AB and EF
9. (b)  $AC \rightarrow DE \rightarrow ABC \rightarrow ABCDE$
10. (a) the left plot represents 3 clusters, the right plot represents 4 clusters,  
(c) the gray cluster (top one) on the left plot has more points than the gray cluster (top one) on the right
11. (a) True
12. (b) 1