

1. (3 points) Which of the following sequences is correct for a k -means algorithm?
 1. Define the number of clusters.
 2. Assign cluster centroids randomly.
 3. Assign each data point to the nearest cluster centroid.
 4. Re-assign each data point to the nearest cluster centroid.
 5. Re-compute the cluster centroids.
 - (a) 1, 2, 3, 5, 4
 - (b) 1, 3, 2, 4, 5
 - (c) 2, 1, 3, 4, 5
 - (d) None of the above.
2. (3 points) For two runs of k -means clustering is expected to get the same clustering results?
 - (a) True
 - (b) False
 - (c) None of the above.
 - (d) It depends.
3. (3 points) Which of the following metrics can we use for finding similarity between two clusters in hierarchical clustering?
 - (a) single linkage
 - (b) average linkage
 - (c) complete linkage
 - (d) All of the above.
4. Consider the `customer_interactions.csv` datafile. This file contains information related to the amount spent by customers on your products, and the number of times they have interacted with your business (for instance, visiting your website). **In Python**, answer the following:
 - (a) (3 points) Using the pandas library, read the csv file and create a data-frame called `interactions`.
 - (b) (4 points) Calculate the z -score standardized values of the two variables and store them in new columns named `z_spend` and `z_interactions`.
 - (c) (4 points) Using the k -means algorithm, cluster the data into four clusters.
 - (d) (4 points) Using hierarchical clustering, cluster the data. Use the average linkage.