1. (1 point) Which of the following is / are true about k-means clustering?

    1. k-means is extremely sensitive to cluster center initialisation
    2. Bad initialization can lead to poor convergence speed
    3. Bad initialization can lead to bad overall clustering
        A. 1 and 2
        B. 1 and 3
        C. 2 and 3
        D. 1,2 and 3

   Answer: D

   Solution: Self explanatory.

2. (1 point) What should be the best choice of no. of clusters based on the following graph:
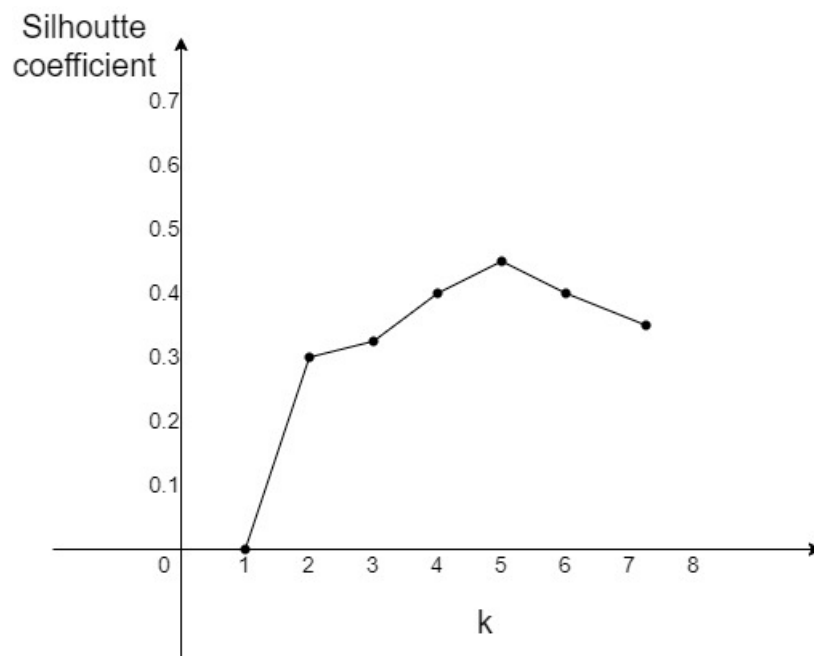

Figure 1: Question no 5

        A. 2
        B. 3
        C. 4

     D. 5

Answer: D

Solution: Since higher the silhouette score better is the clustering.

3. (1 point) Which of the following can act as possible termination conditions in K-Means?

    1. Number of iterations.

    2. Assignment of observations to clusters does not change between iterations.

    3. Centroids do not change between successive iterations.

      A. 1 and 2

      B. 2 and 3

      C. 1 and 3

      D. 1, 2 and 3

Answer: D

Solution: Self explanatory.

4. (1 point) What is the best choice for number of clusters that can be used for k-means clustering based on the following graph?
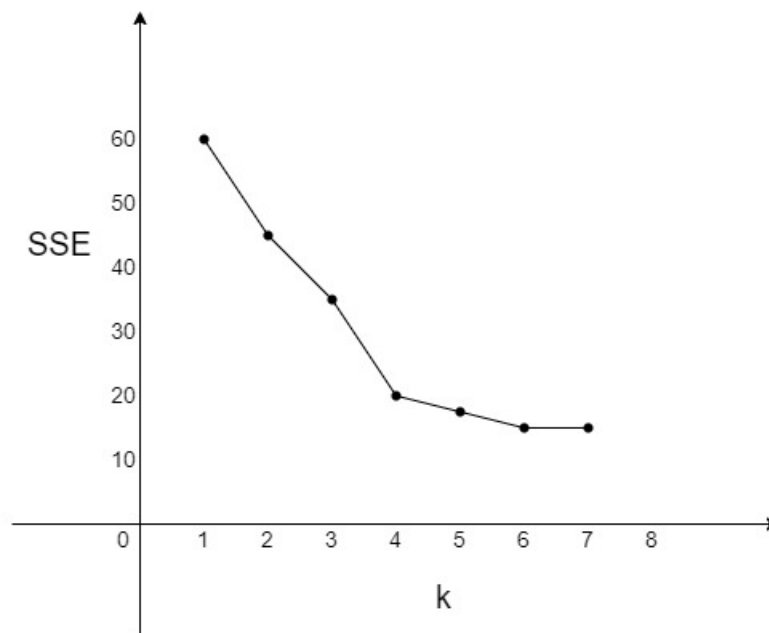

Figure 2: Question no 4

      A. 1

      B. 2

    C. 3

    D. 4

Answer: D

Solution: Look for the elbow like pattern in the chart and identify the elbow point.

5. (1 point) Let there be two features x_1 and x_2 based on which we want to perform k-means clustering. Now consider the following statements

    1. If x_1 and x_2 has a correlation of 1, the cluster centroids will be in a straight line

    2. If x_1 and x_2 has a correlation of 0, the cluster centroids will be in a straight line

    A. 1 only

    B. 2 only

    C. 1 and 2

    D. None

Answer: A

Solution: If the features have correlation of 1, that means the data points form a straight line in 2D plane, hence, the clusters must also be on the same line.

6. (1 point) We want to cluster 9 observations into 3 clusters using K-Means clustering algorithm. After first iteration clusters, C1, C2, C3 has following observations:
C1: (1,1), (1,3), (1,2)
C2: (0,4), (4,0), (5,5)
C3: (2,2), (9,9), (4,4)
What will be the cluster centroids if you want to proceed for second iteration?

    A. C1:(1,1), C2:(4,0), C3:(9,9)

    B. C1:(1,2), C2:(2,2), C3:(7,7)

    C. C1:(1,2), C2:(3,3), C3:(5,5)

    D. C1:(2,1), C2:(4,2), C3:(0,0)

Answer: C

Solution: Take mean of data points present in each cluster

- Centroid of C1 $= \left(\dfrac{1+1+1}{3}, \dfrac{1+3+2}{3} \right) = (1,2)$

- Centroid of C2 $= \left(\dfrac{0+4+5}{3}, \dfrac{4+0+5}{3} \right) = (3,3)$

- Centroid of C3 $= \left(\dfrac{2+9+4}{3}, \dfrac{2+9+4}{3} \right) = (5,5)$