Q1: K-means clustering aims to partition n points into k clusters in which each point belongs to cluster with the nearest__

a) Median
b) Mode
c) Mean
d) None of the above

Q2: Choose the correct statement:

a) K-means++ guarantees good clustering
b) With Elbow Method, as k increases, the loss value monotonically decreases
c) Lloyd K-means is always efficient
d) K-means tends to converge to local minima

Q3: In Elbow Method, we may plot the average cluster radius, what is the correct formula?

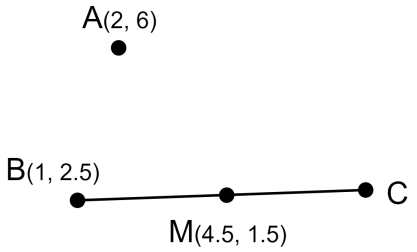
a) $\frac{1}{n}\sum_{i=1}^{k}\sum_{x\in C_i}\|x-c_i\|^2$

b) $\frac{1}{n-k}\sum_{i=1}^{k}\sum_{x\in C_i}\|x-c_i\|^2$

c) $\sum_{i=1}^{k}\sum_{x\in C_i}\|x-c_i\|^2$

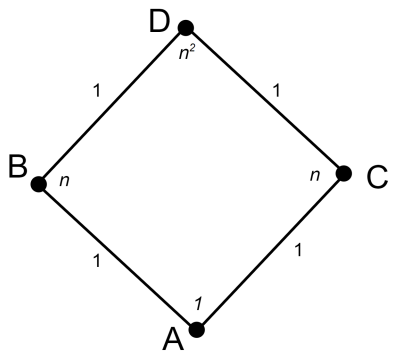
d) $\frac{1}{n}\sum_{i=1}^{k}\sum_{x\in C_i}\|x-c_i\|$

Q4: Given the following figure, find the centroid of 3 points A, B, C, where M is the midpoint of B, C?



a) (3.667, 3)
b) (3, 4)
c) (5, 3)
d) (3, 3.667)

Q5: Given the following dataset. C is the current centroid. Using k-means++, determine the probability A is chosen next.



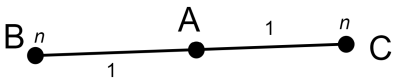
a) $\frac{1}{(n+1)^2}$

b) $\frac{1}{n^2+\sqrt{2}n+1}$

c) $\frac{1}{n^2+2\sqrt{2}n+1}$

d) $\frac{1}{n^2+2n}$

Q6: Given the following dataset. Assume k = 2 and we randomly initiate with centroids A and C. What is the *cost* after the first *Update*?



a) $\frac{2n}{n+1}$

b) $\frac{n}{n+1}$

c) 0

d) 1

Q7: Given the following dataset. k = 2. Using *random initialization*, determine the probability B and A are chosen.

a) $\frac{1}{n+1}$

b) $\frac{n}{(n+1)^4}$

c) $\frac{1}{(n+1)^4}$

d) $\frac{1}{(n+1)^3}$