In K-Means:

(A) **[Ans]** K is often much smaller than N.

(B) K is often larger than N

(C) K is often equal to N.

(D) K is often odd.

(E) None of the above

K-Means:

(A) maximizes the sum of within cluster variances

(B) reaches the same final answer irrespective of the initialization.

(C) **[Ans]** cluster assignments are mutually exclusive and collectively exhaustive.

(D) never converges to the global optima.

(E) None of the above

Consider a measure computed from the final answer of K-Means:

$$J_k = \frac{1}{k} \sum_{l} \sum_{x_i \in C_l} ||x_i - \mu_l||_2^2$$

With increase in k

(A) J_k will monotonically increase

(B) **[Ans]** J_k will monotonically decrease.

(C) It could increase first and then decrease.

(D) It could decrease and then increase

(E) None of the above.

Computational complexity/effort of K-means algorithm depends on:

(A) **[Ans]** K

(B) **[Ans]** N

(C) **[Ans]** d

(D) **[Ans]** No of iterations to converge

(E) **[Ans]** All the above

Consider a set of 10 2D points (i.e., $N = 10, d = 2$) $\{\mathbf{x_i}\}$ as

$$\{[-2,-1]^T,[-3,-2]^T,[0,-1]^T,[-1,0]^T,[2,3]^T,[-1,-2]^T,[3,2]^T,[3,3]^T,[1,1]^T,[2,2]^T\}$$

Cluster them into two clusters $K = 2$. Initialize the K Means such that the first five samples are in cluster A and the next 5 are in cluster B.

Write the final means as

$$(x, y) \text{ and } (a, b)$$