1) What would be the correct relationship among the following three quantities?: ***1 po***

(1) $\sum_{i=1}^{n} ||x_i - \mu_{z_i^t}^t||^2$,

(2) $\sum_{i=1}^{n} ||x_i - \mu_{z_i^{t+1}}^t||^2$ and

(3) $\sum_{i=1}^{n} ||x_i - \mu_{z_i^{t+1}}^{t+1}||^2$

where $\mu_{z_i^t}^t$ and $\mu_{z_i^{t+1}}^{t+1}$ refer to means of cluster z_i in iterations t and $t+1$ respectively. And $\mu_{z_i^{t+1}}^t$ is the mean of the cluster z_i where x_i is going to move in the next (i. e., $(t+1)^{th}$) iteration.

- (1) > (2) < (3)
- (1) < (2) < (3)
- (1) > (2) > (3)
- (1) < (2) > (3)

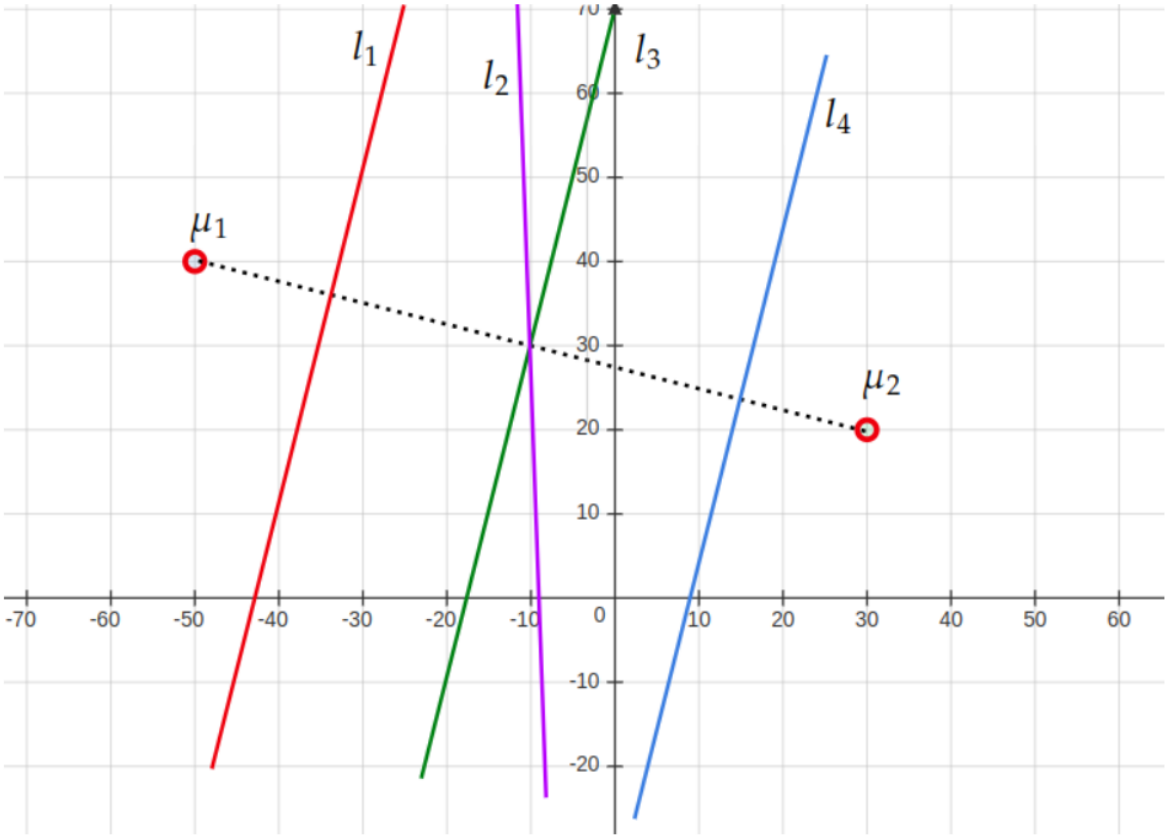
2) Consider that in an iteration t of Lloyd's algorithm, the partition configuration (P^t) is $z_1^t, z_2^t, ..., z_n^t$ where each $z_i^t \in 1, 2, ..., k$. Assume that the algorithm does not converge in iteration t, and hence some re-assignment happens, thus updating the partition configuration in the next iteration (P^{t+1}) to $z_1^{t+1}, z_2^{t+1}, ..., z_n^{t+1}$. How can we say that partition configuration P^{t+1} is better than P^t? ***1 point***

- The value of the objective function for P^{t+1} should be more than that for P^t
- The value of the objective function for P^{t+1} should be lesser than that for P^t
- The value of the objective function for P^{t+1} and P^t should be same.

3) With respect to Lloyd's algorithm, choose the correct statements:

- [x] At the end of k-means, the objective function settles in a local minima and reaching global minima may not be guaranteed.
- [ ] At the end of k-means, the objective function always settles in the global minima.
- [ ] The clusters produced by K-means are optimal.
- [ ] If the resources are limited and the data set is huge, it will be good to prefer K-means over K-means ++.
- [ ] In practice, k should be as large as possible.

4) Consider two cluster centres μ_1 and μ_2 corresponding to two clusters C_1 and C_2 as shown in the below image. Consider four half spaces represented by lines l_1, l_2, l_3 and l_4. Where would the data points falling in cluster C_1 lie?



To the left of l_1

Between l_1 and l_2

Between l_3 and l_4

To the left of l_3

To the left of l_2

5) Which of the following best represents a valid voronoi diagram for K-means algorithm with K = 3? (The dots represent the cluster centres of respective clusters.)

IMAGE-1

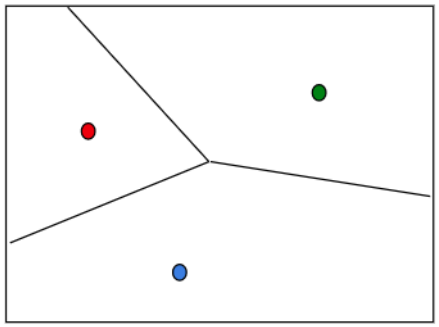


IMAGE-2

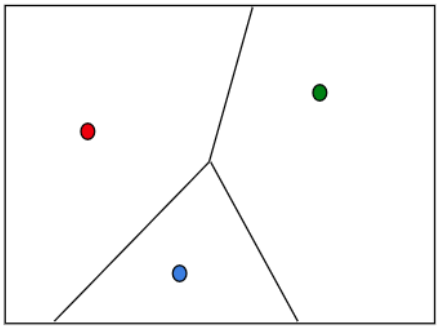


IMAGE-3

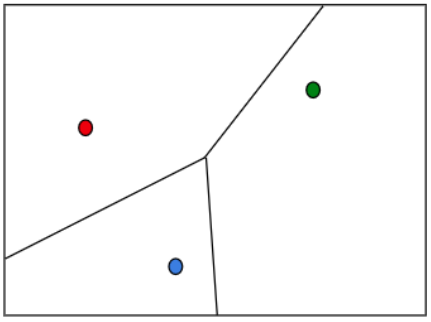
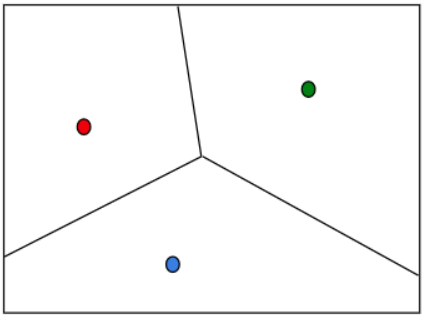
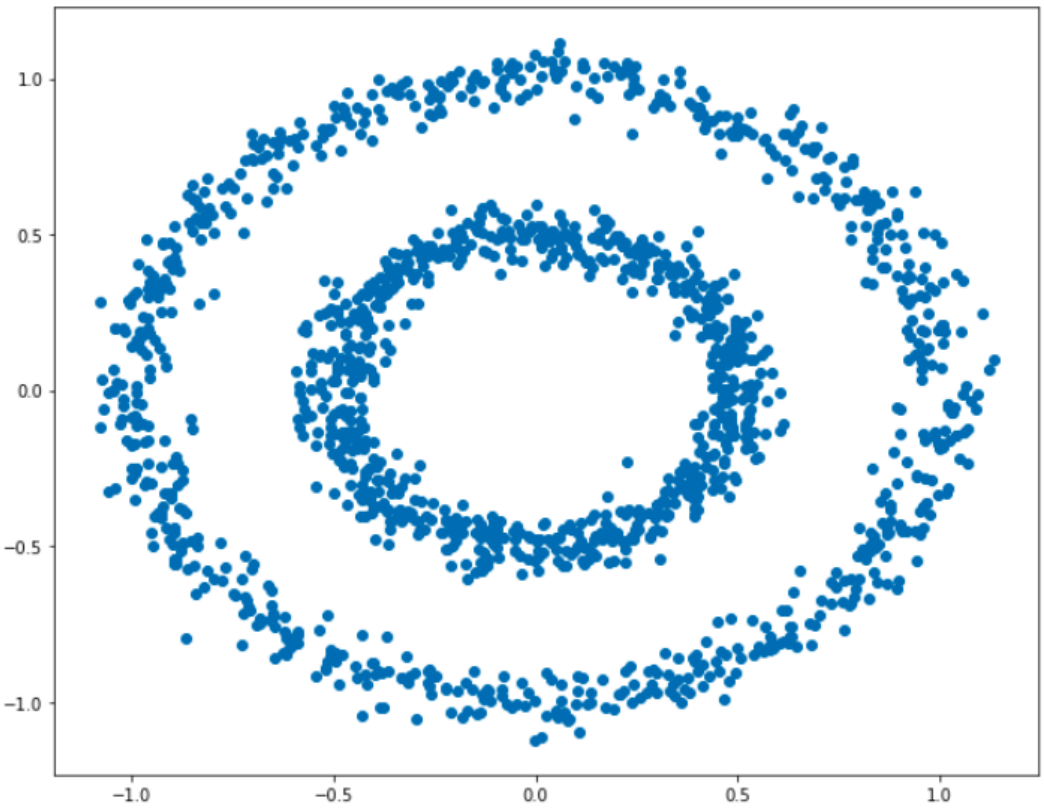


IMAGE-4



6) Consider the following data points:



Assume that K-means is applied on this data with ***k* = 2**. Which of the following are expected to be the clusters produced?

Note: Different colours represent different clusters.

Image-1

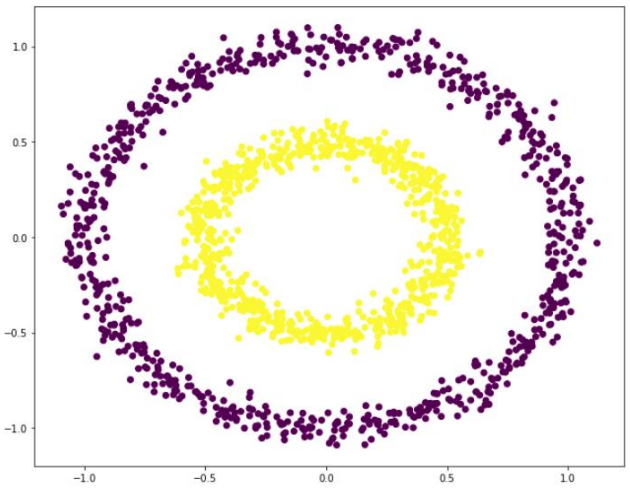
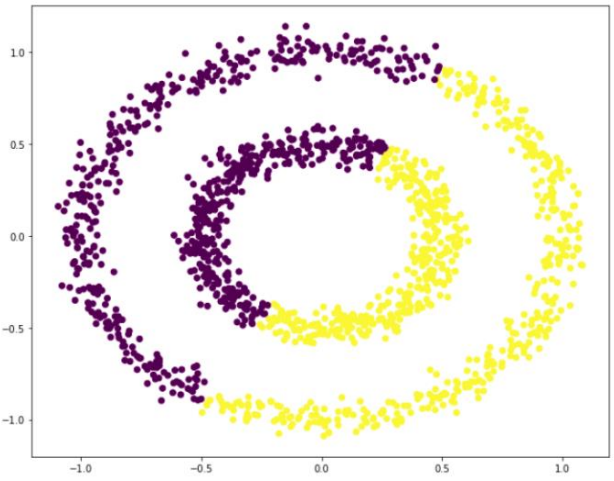


Image-2



7) Assume that in the initialization step of k-means++, the squared distances from the closest mean for 10 points $x_1, x_2, \ldots, x_{10}$ are: 25, 67, 89, 24, 56, 78, 90, 85, 35, 95. Which point has the highest probability of getting chosen as the next mean and how much will that probability be?

- x_4, 0.24
- x_4, 0.037
- x_{10}, 0.95
- x_{10}, 0.1475

8) Consider 7 data points $x_1, x_2, \ldots, x_7$: {(0, 4), (4, 0), (2, 2), (4, 4), (6, 6), (5, 5), (9, 9)}. Assume that we want to form 3 clusters from these points using K-Means algorithm. Assume that after first iteration, clusters $C1, C2, C3$ have the following data points:

$C1$: {(2,2), (4,4), (6,6)}

$C2$: {(0,4), (4,0)}

$C3$: {(5,5), (9,9)}

After second iteration, which of the clusters is the data point (2, 2) expected to move to?

- C_1
- $C2$
- C_3
- Can't say, it is not deterministic.

9) Which of the following statements are True?

K-means is sensitive to cluster center initializations.
Bad initialization can lead to poor convergence speed.
Bad initialization can lead to bad overall clustering.

- 1 and 3
- 1 and 2
- 2 and 3
- 1, 2, and 3

10) If the data set has two features x_1 and x_2, which of the following are true for K means clustering with k = 3?

If x_1 and x_2 have a correlation of 1, the cluster centres will be in a straight line.
If x_1 and x_2 have a correlation of 0, the cluster centres will be in straight line.

- (●) 1
- ( ) 2
- ( ) None of these. Correlation does not affect cluster centres' position.