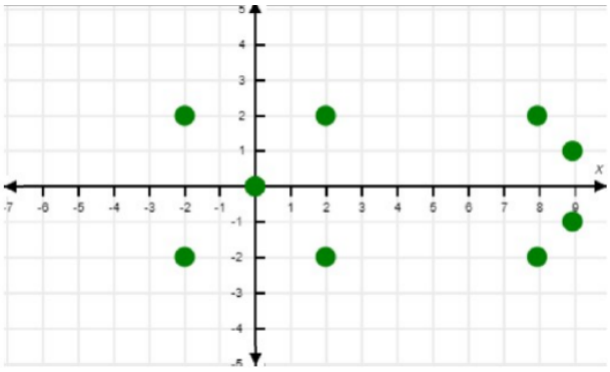
Week_3_mlt_questions

Jan 2024 Quiz 1

1. **Comprehension:**
Consider the data points shown in the following image:



Based on the above data, answer the given subquestions.

i. Perform K-means clustering with $K = 2$ and initial cluster centers at $(0, 0)$ and $(6, 0)$. What are the final means of clusters after convergence?

(a) $(2, 0)$ and $(8, 0)$
(b) $(2, 0)$ and $(8.5, 0)$
(c) $(0, 0)$ and $(8, 0)$
(d) $(0, 0)$ and $(8.5, 0)$
(e) None of these

[Ans: d]

ii. After introducing a new data point $(7, 0)$, the cluster centers were updated. Enter the sum of the updated x-coordinates for both cluster centers accurately, rounding your answer to two decimal places.

[Ans: 8.18 to 8.22]

May 2024 Quiz 1

2. Consider the k-means++ algorithm with $k = 2$ for a dataset that has 7 points: $D = \{x_1, \ldots, x_7\}$. The data-point x_7 has been chosen as the first mean, that is, $\mu_1 = x_7$. The Euclidean distance of the remaining points from this mean is given below:

Point	$d(x_i, \mu_1)$
x_1	1
x_2	2
x_3	$\sqrt{2}$
x_4	$\sqrt{3}$
x_5	$\sqrt{5}$
x_6	1

What is the probability of choosing x_2 as the second mean given that x_7 has been chosen as the first mean?

[Ans: 0.25]

3. **Comprehension:**
Consider the following dataset of six points in $\mathbb{R}^2$:

x	y
-1	-4
5	6
-2	-2
4	5
-3	-3
3	4

K-means clustering is run on this dataset with $k = 2$. In this version, the means are initialized first. The mean of the first cluster is initialized to $(-4, -4)$ and the mean of the second cluster is initialized to $(6, 6)$.
Based on the above data, answer the given subquestions.

i. If (x_1, y_1) is the mean of the first cluster after convergence, find the value of $x_1 + y_1$.

[Ans: -5]

ii. If (x_2, y_2) is the mean of the second cluster after convergence, find the value of $x_2 + y_2$.

[Ans: 9]

Sep 2024 Quiz 1

4. Consider the following data points for k-means clustering.

$$(-1,0),(-1,1),(-1,-1),(2,0),(3,1),(3,-1),(4,0)$$

In the initialization step of k-means with $k=2$, suppose $\mu_1^0=(-1,0)$ and $\mu_2^0=(2,0)$. Distances of datapoints from initial cluster means is tabulated below:

x_i	$\|x_i-\mu_1^0\|_2^2$	$\|x_i-\mu_2^0\|_2^2$
$(-1,0)$	0	3
$(-1,1)$	1	10
$(-1,-1)$	1	10
$(2,0)$	3	0
$(3,1)$	17	2
$(3,-1)$	17	2
$(4,0)$	5	2

As per these cluster centers, the data points are then assigned to either cluster 1 or cluster 2. After this assignment, what will be the value of the objective function?
Note: Objective function is given by

$$F(z_1,z_2,\ldots,z_n)=\sum_{i=1}^{n}\|x_i-\mu_{z_i}\|_2^2$$

[Ans: 6]

5. **Comprehension:**
A k-means++ algorithm with $k=3$ is applied on the following 2D points:

$$(0,1),(1,0),(1,2),(2,1),(2,3),(2,4),(3,2)$$

First cluster mean μ_1^0 is chosen as $(2,1)$.
Suppose the point with the highest score is chosen as the 2nd cluster mean μ_2^0.
Based on the above data, answer the given subquestions.

i. What is μ_2^0? Use squared distance to calculate the scores.

(a) $(0,1)$
(b) $(2,3)$
(c) $(3,2)$
(d) $(2,4)$

[Ans: d]

ii. Which point has the lowest probability of being chosen as the 3rd cluster mean? Use squared distance to calculate the scores.

(a) $(1,0)$
(b) $(2,3)$
(c) $(3,2)$
(d) $(1,2)$

[Ans: b]

Jan 2025 Quiz 1

6. For the Lloyd's algorithm, what would be the correct relationship among the following three quantities?

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

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

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

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

(a) (1) > (2) > (3)

(b) (1) < (2) < (3)

(c) (3) < (1) < (2)

(d) (2) < (3) < (1)

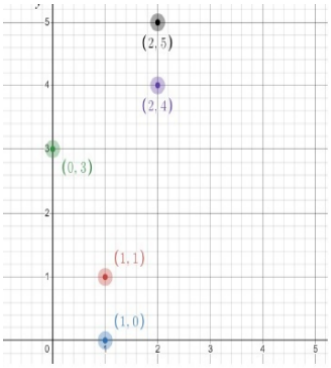
[Ans: c]

7. K-means clustering is run on a dataset of 100 points with $K = 3$. If you know that the points $\begin{bmatrix}0\\1\end{bmatrix}$ and $\begin{bmatrix}1\\1\end{bmatrix}$ are a part of the same cluster $C1$, then which of the following points will definitely lie in $C1$?

(a) $\begin{bmatrix}-1\\1\end{bmatrix}$ (b) $\begin{bmatrix}1\\2\end{bmatrix}$ (c) $\begin{bmatrix}0.2\\1\end{bmatrix}$ (d) $\begin{bmatrix}0\\0\end{bmatrix}$

[Ans: c]

8. K-means++ algorithm was run on the following dataset with $K = 2$.



If we use the point $(1, 1)$ as the first cluster mean, then find the probability of the point that has the least chance of being chosen as the second cluster mean. Enter the answer correct to two decimal places. Use Euclidean distance to calculate the distance.

[Ans: 0.07 to 0.11]

May 2025 Quiz 1

9. How does K-means++ algorithm enhance the initialization process as compared to the standard K-means algorithm?

(a) It ensures the algorithm converges in a fixed number of iterations.

(b) It improves centroid initialization by spreading them out, leading to better separated clusters.

(c) It automatically calculates the best value of K.

(d) It removes the need to run the algorithm multiple times.

[Ans: b]

10. In a K-means algorithm with $K = 3$, the objective value $F(z_1, \ldots, z_n) = \sum_{i=1}^{n} \|x_i - \mu_{z_i}\|^2$ is recorded after each iteration as follows:

Iteration 1: 800, Iteration 2: 500, Iteration 3: 400, Iteration 4: 400, Iteration 5: 400.

At which iteration is the algorithm most likely to reach convergence?

(a) 2

(b) 3

(c) 5

(d) 4

[Ans: b]

11. Which of the following statements is/are true? (MSQ)

(a) EM algorithm converges to a global optimum.

(b) Lloyd's algorithm converges to a local optimum.

(c) Poor initialization does not affect the clusters produced by EM algorithm.

(d) Poor initialization leads to a slower convergence of Lloyd's algorithm.

(e) Lloyd's algorithm is sensitive to the number of clusters (k).

[Ans: b, d, e]

2024 Apr 28 ET

12. Imagine a dataset characterized by two features, Feature 1 and Feature 2, demonstrating a perfect negative correlation of -1. When applying k-means clustering with $k = 3$ to this dataset, what is the most likely arrangement of cluster centers that minimizes the within-cluster sum of squares (WCSS)?

(a) An equilateral triangle centered around the mean of the data.

(b) Cluster centers positioned along a straight line.

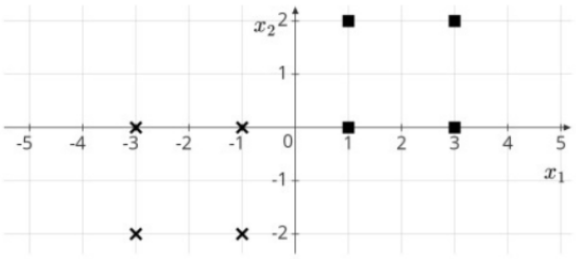
(c) A triangle with two acute angles, positioned strategically within the datadistribution.

(d) A right-angled triangle with one center at the origin.

[Ans: b]

2024 Sep 01 ET

13. The result of k-means clustering on a dataset of eight points is displayed below. Four points belong to the cluster denoted by the ■ symbol and the rest belong to the cluster denoted by the × symbol. The cluster boundary is a line such that all points on it could belong to either of the two clusters. Which of the following is the equation of the cluster boundary?



(a) $2x_1 + x_2 = 0$

(b) $x_1 - 2x_2 = 0$

(c) $x_1 + x_2 = 0$

(d) $x_1 + 2x_2 = 0$

[Ans: a]

2024 Dec 22 ET

14. Which of the following sequences is correct for K-Means algorithm?
    1. Assign each data point to the nearest cluster centers.
    2. Re-assign each point to the nearest cluster centers.
    3. Assign cluster centers randomly.
    4. Re-compute cluster centres.
    5. Specify the number of clusters.

(a) 3, 5, 1, 4, 2

(b) 5, 3, 1, 2, 4

(c) 3, 5, 2, 4, 1

(d) 5, 3, 1, 4, 2

(e) None of these

[Ans: d]

2025 Apr 13 ET

15. Consider the following dataset in $\mathbb{R}^2$:

$$\{(1,2),(3,3),(3,4),(5,4),(7,8),(9,10)\}$$

Suppose we want to form 3 clusters from these points using Lloyd's algorithm. After the first iteration, clusters C_1, C_2, and C_3 have the following data points:

$$C_1 : \{(1,2),(3,3)\}$$
$$C_2 : \{(3,4),(5,4)\}$$
$$C_3 : \{(7,8),(9,10)\}$$

After second iteration, which of the clusters is the data point $(3,3)$ expected to move to?

(a) C_1

(b) C_2

(c) C_3

[Ans: a]

2025 Aug 31 ET

16. Consider K-means clustering algorithm with $K = 2$ on a one-dimensional dataset $D = \{x_1, x_2, \ldots, x_n\}$ aiming to minimize:

$$F(z_1, z_2, \ldots, z_n) = \sum_{i=1}^{n} \|x_i - \mu_{z_i}\|^2,$$

where $z_i \in \{1, 2\}$ is the cluster assignment for point x_i and μ_{z_i} is the mean of the assigned cluster.

For greedy initialization, choose the first center uniformly at random from D, and the second as the point farthest from it (by Euclidean distance). Let F_{optimal} be the minimum possible value of F. Which of the following statements about greedy initialization is true?

(a) Greedy initialization always achieves $F = F_{\text{optimal}}$ by maximizing the distance between initial centers.

(b) Greedy initialization may achieve $F = F_{\text{optimal}}$ or a larger value depending on the dataset and the initial choice.

(c) Greedy initialization does not ensure convergence for K-means clustering.

(d) Greedy initialization ensures convergence to a unique solution for any dataset.

[Ans: b]

2025 Dec 18 ET

17. **Comprehension:**
Consider the following points on the real line: $x_1 = 1, x_2 = 4, x_3 = 7, x_4 = 8$. We run k-means algorithm with $k = 2$. The initial cluster centers are $\mu_1^0 = x_1$ and $\mu_2^0 = x_2$. Based on the above data, answer the given subquestions.

i. What will be the cluster means after iteration 1?

(a) $\mu_1^{(1)} = 2, \mu_2^{(2)} = 6$

(b) $\mu_1^{(1)} = 1, \mu_2^{(2)} = 7.5$

(c) $\mu_1^{(1)} = 4, \mu_2^{(2)} = 7$

(d) $\mu_1^{(1)} = 1, \mu_2^{(2)} = 6.33$

[Ans: d]

2026 Sep 13 ET

18. A Euclidean-distance-based k-means clustering algorithm was run on a dataset of 100 points with $k = 3$. If the points $(2, 2)$ and $(2, -2)$ are both part of cluster 3, then which ONE of the following points is necessarily also part of cluster 3?

(a) $(0, 0)$

(b) $(4, 0)$

(c) $(2, 0)$

(d) $(0, 2)$

[Ans: c]

Oct 2022 Quiz 1

19. Consider the scenario where you observe that 8 of your friends $\{f_1, \ldots, f_8\}$ have scored the following marks in a quiz: $\{6, 3, -5, -4, 2, -3, 5, -2\}$, respectively. Which of the cluster initializations given below will result in clusters where your friends with positive marks are in one cluster and the rest are in another after executing one step of Lloyd's algorithm?

$$I_1 : z_1 = z_2 = z_3 = z_4 = z_5 = z_6 = z_7 = 2, \quad z_8 = 1,$$
$$I_2 : z_1 = z_2 = z_3 = z_4 = 1, \quad z_5 = z_6 = z_7 = z_8 = 2.$$

(a) I_1: No, I_2: No

(b) I_1: Yes, I_2: Yes

(c) I_1: Yes, I_2: No

(d) I_1: No, I_2: Yes

(e) Insufficient information

[Ans: b]

20. Consider the following dataset:

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 0 \\ -2 \end{bmatrix},$$
$$\mathbf{x}_5 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \mathbf{x}_6 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_7 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

For $k = 3$, assume that the following cluster assignment is given to us, which shows the clusters assigned to all data points except $\mathbf{x}_3$:

$$\mathbf{z} = \{3, 1, ?, 2, 1, 3, 2\}.$$

What should be the cluster assigned to the data point $\mathbf{x}_3$?

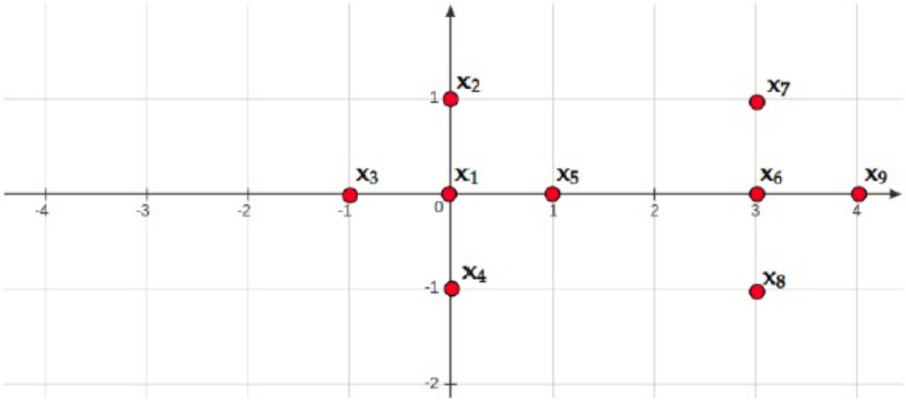
(a) 1

(b) 2

(c) 3

[Ans: a]

21. Consider the data points shown in the following image:



Consider that in the initialization step of K-means with $k = 2$, the data points $\mathbf{x}_1$ and $\mathbf{x}_6$ got selected as initial cluster centers. That is, $\mu_1^0 = (0, 0)$ and $\mu_2^0 = (3, 0)$. As per these cluster centers, the data points were then assigned to either cluster 1 or cluster 2. After this assignment, what will be the value of the objective function (Obj)? Further, as per the assignment of data points, the cluster centers will be recomputed for the next iteration. What will be the values of μ_1^1 and μ_2^1?

(a) Obj $= 7$, $\mu_1^1 = (1, 0)$, $\mu_2^1 = (3.33, 0)$

(b) Obj $= 17$, $\mu_1^1 = (1, 0)$, $\mu_2^1 = (3.25, 0)$

(c) Obj $= 7$, $\mu_1^1 = (0, 0)$, $\mu_2^1 = (3.25, 0)$

(d) Obj $= 17$, $\mu_1^1 = (0, 0)$, $\mu_2^1 = (3.33, 0)$

[Ans: c]

22. **Comprehension:**
Consider the following dataset:

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\mathbf{x}_4 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \quad \mathbf{x}_5 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \mathbf{x}_6 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}.$$

With $k = 2$, as per k-means++, what is the probability of the following points (in the given order) being chosen as initial cluster centers? Enter your answers correct to three decimal places.

i. What is the probability of the points $\mathbf{x}_2, \mathbf{x}_1$ (in that order) being chosen as initial cluster centers?

[Ans: 0.033]

ii. What is the probability of the points $\mathbf{x}_2, \mathbf{x}_3$ (in that order) being chosen as initial cluster centers?

[Ans: 0.017]

iii. What is the probability of the points $\mathbf{x}_2, \mathbf{x}_5$ (in that order) being chosen as initial cluster centers?

[Ans: 0.067]

Jul 2023 Quiz 1

23. Consider the following hard K-means clustering problem with four points: $P(1,1)$, $Q(2,1)$, $R(4,3)$, and $S(5,4)$. Consider the number of clusters to be $k=2$ and the initial centroids to be $C_1=(0,0)$ and $C_2=(4,4)$. After how many iterations will the algorithm terminate and what will be the final centroids? Use the Euclidean distance metric to calculate distance.

(a) 2, $(1,1)$, $(4,4)$

(b) 3, $(1,1)$, $(4,4)$

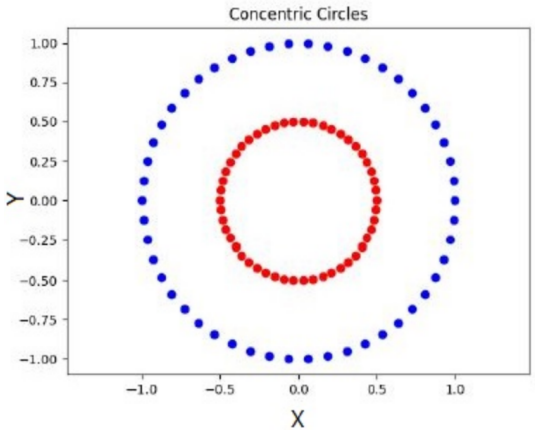
(c) 2, $(1.5,1)$, $(4.5,3.5)$

(d) 4, $(1.5,1)$, $(4.5,3.5)$

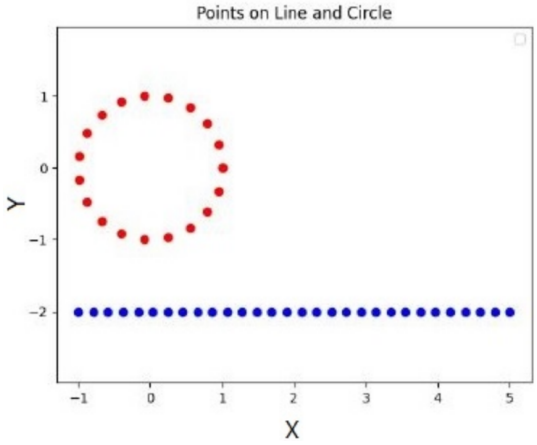
[Ans: c]

24. Which of the following clusters cannot be obtained as final clusters using Lloyd's algorithm? Assume that Lloyd's algorithm was run on these datasets with $k=2$. (MSQ)

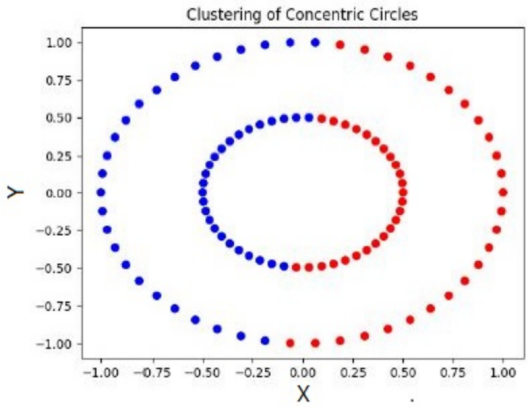
(a)



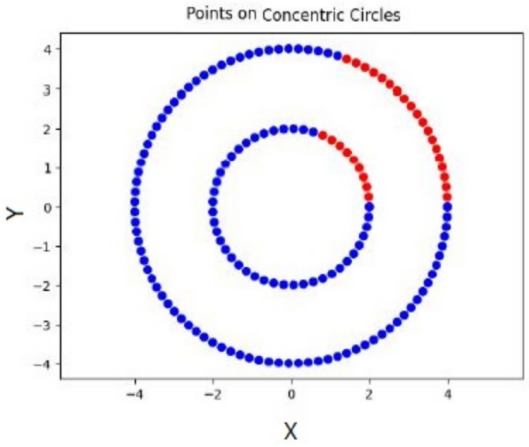
(b)



(c)



(d)



[Ans: a, b, d]

25. Consider the following dataset in $\mathbb{R}$:

$$\{-3, -1, 0, 1, 4, 299, 300, 301\}.$$

Lloyd's algorithm ($K = 2$) is run on this dataset with the points -10 and 310 as the initial cluster centers. Let a, b be the final cluster centers on convergence. What is the value of the product ab?

[Ans: 60]

Oct 2023 Quiz 1

26. In the K-means algorithm, cluster centroids (μ_k) are updated at each iteration to minimize the sum of squared distances between data points and their assigned centroids. The formula used for this update is:

$$\mu_k = \frac{\sum_i \mathbf{x}_i \mathbf{1}(z_i = k)}{\sum_i \mathbf{1}(z_i = k)}.$$

Here $\mathbf{x}_i$ represents the i-th data point in the dataset, and z_i represents the cluster assignment for $\mathbf{x}_i$, where z_i is an integer indicating the cluster index k.

Suppose you have a dataset of 9 two-dimensional data points and are performing K-means clustering with $K = 3$ clusters. After the first iteration, clusters C_1, C_2, C_3 have the following data points:

$$C_1 : \{(2, 2), (3, 3), (4, 4)\},$$
$$C_2 : \{(6, 6), (7, 7), (8, 8)\},$$
$$C_3 : \{(10, 10), (11, 11), (12, 12)\}.$$

After the second iteration, which of the following options represents the updated cluster centroid μ_k for C_1 using the provided formula?

(a) $\mu_1 = (2.5, 2.5)$

(b) $\mu_1 = (4.5, 4.5)$

(c) $\mu_1 = (3, 3)$

(d) $\mu_1 = (5.5, 5.5)$

[Ans: c]

27. Consider 9 data points $\mathbf{x}_1, \ldots, \mathbf{x}_9$:

$$(2,3), (1,2), (3,1), (5,6), (7,8),$$
$$(9,9), (6,7), (10,10), (11,11).$$

Suppose we aim to form 3 clusters from these points using the K-means algorithm. After the first iteration, clusters C_1, C_2, C_3 have the following data points:

$$C_1 : \{(1,2), (3,1)\},$$
$$C_2 : \{(2,3), (5,6), (6,7)\},$$
$$C_3 : \{(9,9), (7,8), (10,10), (11,11)\}.$$

After the second iteration, to which cluster will the data point $(2,3)$ be reassigned?

(a) C_1

(b) C_2

(c) C_3

(d) It remains in the same cluster as in the first iteration.

[Ans: a]

28. **Comprehension:**
A k-means++ algorithm with $k = 3$ was applied to the following 2D points:

$$(2,2), (3,3), (4,4), (10,10), (12,12), (13,13).$$

$(2,2)$ is chosen as the first cluster mean. Based on the above data, answer the given subquestions.

i. Which point has the highest probability of being chosen as the second cluster mean? Use Manhattan distance to compute the distances.

(a) $(4,4)$

(b) $(12,12)$

(c) $(10,10)$

(d) $(13,13)$

[Ans: d]

ii. Assume that $(3,3)$, $(10,10)$, and $(12,12)$ are the cluster means. You introduce a new data point $(-5,10)$. Which cluster mean is the data point $(-5,10)$ closest to when using the Euclidean distance method?

(a) $(3,3)$

(b) $(10,10)$

(c) $(12,12)$

(d) None of these

[Ans: a]

Source note: The supplied reference corrects the quizpractice.space answer keys for both subquestions.

Feb 2024 Quiz 1

29. Consider Lloyd's algorithm used for k-means clustering and choose the correct statements. (MSQ)

(a) K-means algorithm may get stuck at local minima.

(b) It guarantees finding the optimal clustering (global minimum) in every run.

(c) In practice, k should be as large as possible.

(d) If the resources are limited and the dataset is huge, it will be good to prefer K-means over K-means++.

[Ans: a, d]

Oct 2025 Quiz 1

30. Which among the following is true with respect to Lloyd's algorithm? (MSQ)

(a) It always converges.

(b) It may fail to converge for an initialization while it may converge for another initialization.

(c) It is sensitive to outliers.

(d) The solution produced is always the optimal.

[Ans: a, c]

31. Consider the dataset consisting of the following points:

$$(0,0), (0,2), (2,0), (2,2).$$

We want to run the K-means++ algorithm with $K = 2$. Suppose the first cluster center chosen is $(0,0)$. Which of the following statements are correct? (MSQ)
Hint: The score of a data point $\mathbf{x}$ is given by

$$S(\mathbf{x}) = \min_{j=1,2,\ldots,l-1} \|\mathbf{x} - \mu_j^0\|^2,$$

where μ_j^0 is the initial cluster center.

(a) The points $(0,2)$, $(2,0)$, and $(2,2)$ have equal probability of being chosen as the second cluster center.

(b) The probability of selecting $(2,2)$ as the second cluster center is twice that of selecting $(0,2)$.

(c) If $(2,2)$ is chosen as the second cluster center, the algorithm will always converge in one iteration.

(d) The tie-breaking rule will affect the final cluster centers.

[Ans: b, c, d]

32. **Comprehension:**
Consider the following data points shown in Figure 2. Based on the above data, answer the given subquestions.

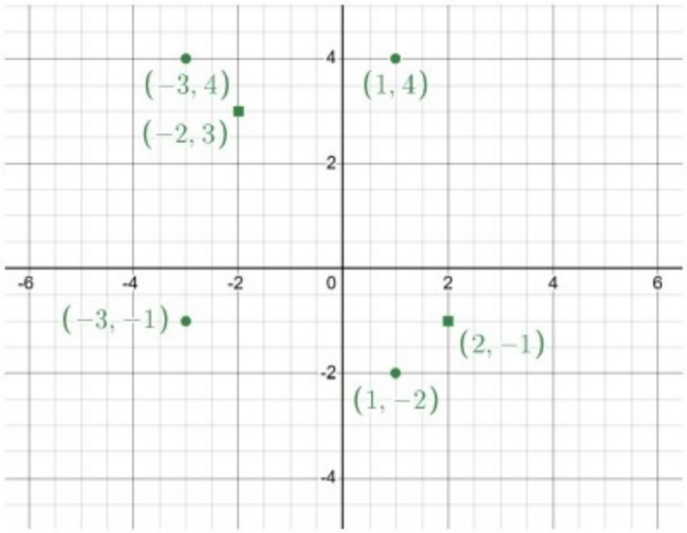


Figure 2: Dataset 2

i. Perform K-means clustering with $K = 2$ and initial cluster centers at $(-2, 3)$ and $(2, -1)$. What are the final means of clusters after convergence?

(a) $(-1.75, 2.5)$ and $(1.5, -1.5)$
(b) $(-2, 3)$ and $(2, -1)$
(c) $(-1.33, 3.66)$ and $(0, -1.33)$
(d) $(1.33, 1)$ and $(2.66, 2)$

[Ans: a]

ii. Suppose we introduce a new data point $(0, 0)$ to the existing dataset and perform K-means with the same initial cluster centers. How many iterations would it require now for convergence?

[Ans: 1]

Mar 2026 Quiz 1

33. Consider a clustering problem with 100 data points and two clusters ($k = 2$). As initialization, a subset of the dataset comprising the first fifty data points, call this D_1, are assigned to cluster 1 and the remaining points, call this D_2, are assigned to cluster 2. K-means is run for this configuration and it converges after 10 iterations. Consider the following scenarios:
Scenario 1: After convergence, it is observed that all points in D_1 are assigned to cluster 2 and all points in D_2 are assigned to cluster 1.
Scenario 2: After convergence, it is observed that all points in D_1 are assigned to cluster 1 and all points in D_2 are assigned to cluster 2.

Which of the following is true?

(a) Scenario 1 is possible, but scenario 2 is impossible.

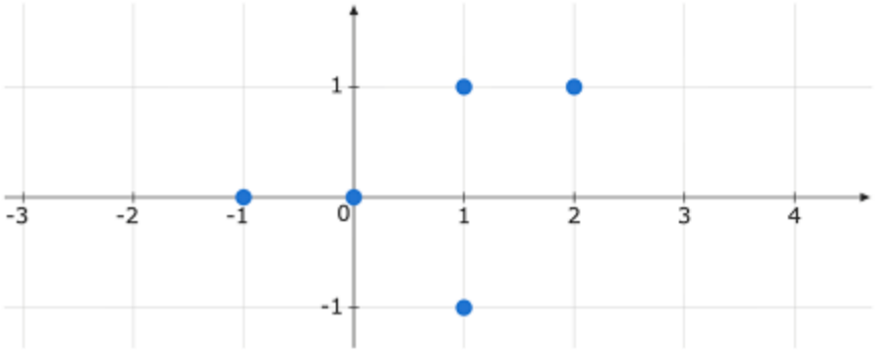
(b) Scenario 1 is impossible, but scenario 2 is possible.

(c) Both scenarios are possible.

(d) Both scenarios are impossible.

[Ans: d]

34. Consider the following data points in a clustering problem with five data points and three clusters ($k = 3$):



K-means++ is used to initialize the three means. The first mean is chosen to be $\mu_1 = (0, 0)$ and the second mean to be $\mu_2 = (1, 1)$. Find the probability of choosing $(1, -1)$.

[Ans: 0.5]

35. **Comprehension:**
X is a data matrix in $\mathbb{R}^2$ for a clustering problem with two clusters ($k = 2$). The i-th data point is represented as (x_i, y_i).

i	x_i	y_i
1	0	0
2	3	0
3	4	1
4	1	1

$\mathbf{z}^{(t)} = (z_1^{(t)}, z_2^{(t)}, z_3^{(t)}, z_4^{(t)})$ indicates the cluster assignment vector at the end of time step t. $\mathbf{z}^{(0)}$ indicates the initial assignment. The cluster indices are represented by 1 and 2, that is, $z_i^{(t)} \in \{1, 2\}$. Based on the above data, answer the given subquestions.

i. If $\mathbf{z}^{(0)} = (1, 1, 2, 2)$, which of the following is the final cluster assignment after running K-means on this dataset with this initial configuration?

(a) $(1, 1, 2, 2)$
(b) $(1, 2, 2, 1)$
(c) $(1, 2, 1, 2)$
(d) $(1, 1, 1, 2)$
(e) $(1, 1, 2, 1)$

[Ans: b]

ii. If $\mathbf{z}^{(0)} = (1, 2, 2, 1)$, which of the following is the final cluster assignment after running K-means on this dataset with this initial configuration?

(a) $(1, 1, 2, 2)$
(b) $(1, 2, 2, 1)$
(c) $(1, 2, 1, 2)$
(d) $(1, 1, 1, 2)$
(e) $(1, 1, 2, 1)$

[Ans: b]

Jul 2026 Quiz 1

36. Consider the one-dimensional dataset $x_1 = 0$, $x_2 = 2$, $x_3 = 6$, $x_4 = 8$. Suppose K-means is run with $K = 2$ clusters. At iteration t, the centroids are $\mu_1^t = 0$, $\mu_2^t = 10$. The assignment step is performed using these centroids to produce new assignments z_i^{t+1} for each point. Compute

$$\sum_{i=1}^{4} \left\| x_i - \mu_{z_i^{t+1}}^t \right\|^2.$$

(a) 4
(b) 8
(c) 24
(d) 44

[Ans: c]

37. Which of the following statements is always true for the K-means algorithm with $K < n$ cluster centers, between consecutive iterations t and $t+1$?

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

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

(c) If a data point changes its cluster assignment from cluster j to cluster k, it is guaranteed that the centroid of cluster j strictly changes, i.e., $\mu_j^{t+1} \neq \mu_j^t$.

(d) It is not possible for the K-means algorithm to revisit a configuration, where a configuration corresponds to the K-way partition generated by clustering at the end of each iteration.

[Ans: d]

38. A dataset containing 1000 points is clustered using the k-means algorithm. Which of the following values of k is most likely to yield the minimum value of the objective function of k-means?

(a) $k = 1$

(b) $k = 10$

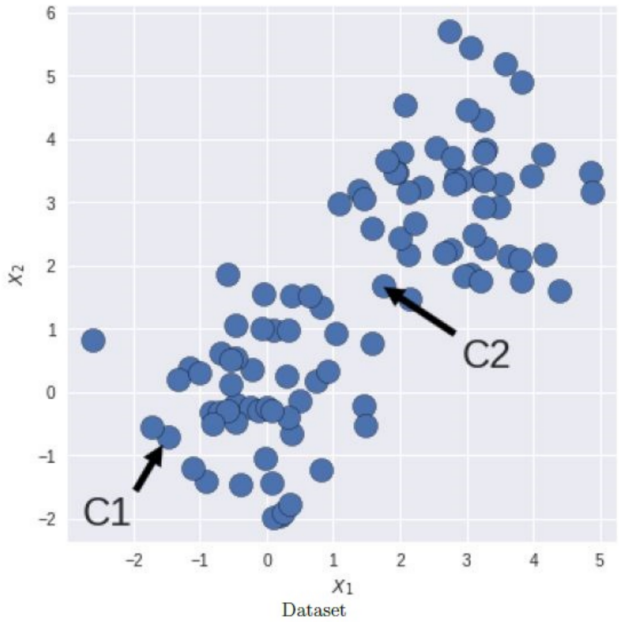
(c) $k = 100$

(d) $k = 1000$

[Ans: d]

Aug 2022 End Term

39. Consider unlabeled data with two features X_1 and X_2 as shown in the figure.



C_1 and C_2 are the coordinates of centroids obtained after certain iterations of the K-means algorithm with $K = 2$. Which of the following options is correct about the silhouette score S? Note: Euclidean distance is used to calculate the distances.

(a) S will be positive and close to 0.

(b) S will be positive and close to 1.

(c) S will be exactly -1.

(d) S will be negative but need not be exactly -1.

[Ans: d]

40. Consider the following data points:

$$(1, 1), (1, 2), (5, 4), (2.4, 3), (2.6, 3).$$

We perform k-means clustering on the above data with $k = 2$, using Manhattan distance as the distance measure. At the t-th iteration, we have $(1, 1)$ as the centroid for cluster 1 and $(5, 4)$ as the centroid for cluster 2. After performing the $(t + 1)$-th iteration, the point $(2.4, 3)$ will belong to:

(a) Cluster 1

(b) Cluster 2

(c) Cannot be determined

[Ans: b]

Dec 2022 End Term

41. Consider that we introduce negative marking in this exam. After getting the results, you observe that eight of your friends $\{f_1, \ldots, f_8\}$ have scored the following marks, respectively:

$$\{5, 6, -2, -3, 1, 7, -4, -1\}.$$

You want to cluster your friends into two groups based on their marks using Lloyd's algorithm. You initialize the algorithm by keeping the first four friends, i.e., $\{f_1, f_2, f_3, f_4\}$, in cluster 1 (C_1) and the last four friends, i.e., $\{f_5, f_6, f_7, f_8\}$, in cluster 2 (C_2). How would the clusters look after executing one step of Lloyd's algorithm?

(a) $C_1 : (f_1, f_2, f_3, f_4)$, $C_2 : (f_5, f_6, f_7, f_8)$

(b) $C_1 : (f_1, f_2, f_5, f_6)$, $C_2 : (f_3, f_4, f_7, f_8)$

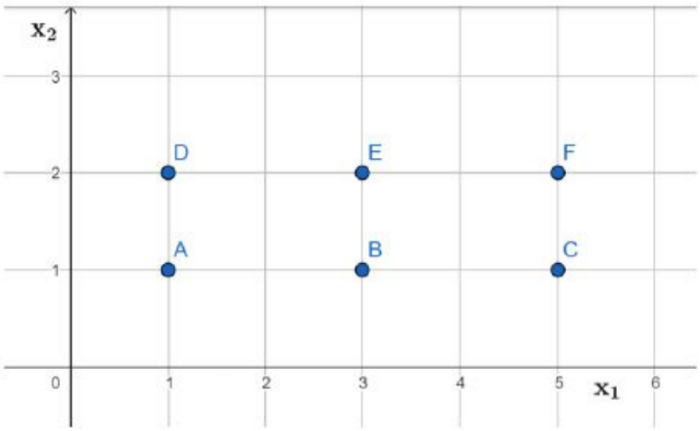
(c) $C_1 : (f_1, f_2, f_6)$, $C_2 : (f_3, f_4, f_5, f_7, f_8)$

(d) $C_1 : (f_2, f_4, f_6)$, $C_2 : (f_1, f_3, f_5, f_7, f_8)$

[Ans: c]

Apr 2023 End Term

42. Consider that the K-means algorithm was run on the following two-dimensional dataset with $K = 3$. The initial cluster centroids are points A, B, and D as shown in the figure.



How many cluster centroids in the final clusters will be the same as the initial cluster centroids? Use Euclidean distance to calculate the distance.

[Ans: 2]

Sep 2023 End Term

43. A clustering of 100 data points in $\mathbb{R}^2$ was done using Lloyd's algorithm with $K = 3$. You are told that the points $[1, 1]^T$, $[-1, -1]^T$, $[1, -1]^T$, and $[-1, 1]^T$ are in the same cluster. Which of the following points will definitely lie in the same cluster? (If there are any ties with this cluster, they should be assigned exclusively to this cluster.) (MSQ)

(a) $[0, 0]^T$

(b) $[2, 2]^T$

(c) $[-0.5, 0.5]^T$

(d) $[0.75, 0.9]^T$

[Ans: a, c, d]

Dec 2023 End Term

44. Given the dataset $\{2, 5, 8, 11, 15, 18, 22, 25\}$, apply Lloyd's algorithm with $k = 3$ and initial cluster centers at 5, 15, and 25. What is the total sum of the final cluster centers when the algorithm converges? Enter your answer correct to two decimal places.

[Ans: 43.17]

Dec 2025 End Term

45. Consider the following points on the real line: $x_1 = 1$, $x_2 = 4$, $x_3 = 7$, $x_4 = 8$. We run the k-means algorithm with $k = 2$. The initial cluster centers are $\mu_1^0 = x_1$ and $\mu_2^0 = x_2$. Find the value of the objective function

$$F(z_1, z_2, z_3, z_4) = \sum_{i=1}^{4} \|x_i - \mu_{z_i}\|^2$$

after convergence. Enter the answer correct to two decimal places.

[Ans: 8.67]