

Questions based on <https://youtu.be/!Un8k5zSl6g?si=6RxTC5LA91CpZEGu>.

Exercise 1

Given the following 2D points, simulate 1 iteration of K-means with $k = 2$:

A = (2, 3)

B = (3, 4)

C = (6, 8)

D = (7, 9)

E = (8, 10)

Initial centroids:

Centroid 1 (Red): (3, 3)

Centroid 2 (Blue): (8, 8)

Exercise 2

For each point (A–E), calculate its distance to both centroids. Assign it to the closest centroid. (Use Euclidean distance; rounding is fine.)

Point	Distance to Red (3,3)	Distance to Blue (8,8)	Assigned to
A			
B			
C			
D			
E			

Exercise 3

Calculate the new centroids by averaging the x and y coordinates of the points assigned to each cluster.

Example: If Centroid 1 is assigned points A and B, new Centroid 1 = (average of A and B).

Questions based on <https://youtu.be/4b5d3muPQmA?si=9kHpiS1M15gteCRZ>.

Exercise 1

Given these 1D data points: **2, 4, 7, 10, 12**, and **K = 2**:

- a) **Initialization**: Randomly pick 2 initial centroids (e.g., 4 and 12).
- b) **Assignment**: Calculate distances and assign each point to the nearest centroid.
- c) **Update**: Compute the mean of each cluster and update centroids.
- d) **Iterate**: Repeat the assignment and update steps until cluster assignments don't change.

Task: Show your work for two full iterations. Which data points end up in each cluster? What are the final centroids?

Exercise 2

You're given the total within-cluster variation (sum of squared distances to centroids) for K values from 1 to 5:

K	Total Variation
1	150
2	80
3	55
4	45
5	40

- a) Sketch the elbow plot.
- b) Which K would you choose based on the “elbow” rule, and why?

Exercise 3

Write concise responses (1–2 sentences) to the following:

1. Why does K-means require choosing K ahead of time?
2. What effect does the random initialization of centroids have on the outcome?
3. When using 2D data or heatmap data, what distance metric does K-means typically use?

Exercise 4

Choose a real dataset (it could be a simple CSV with 2 numeric columns).

1. Run K-means with different values of K (from 2 to 5).
2. Produce scatter plots showing cluster assignments and label centroids.
3. Calculate and plot an elbow graph.

Discussion: What insights do the clusters reveal? Does the elbow plot support your choice of K?