Week 3

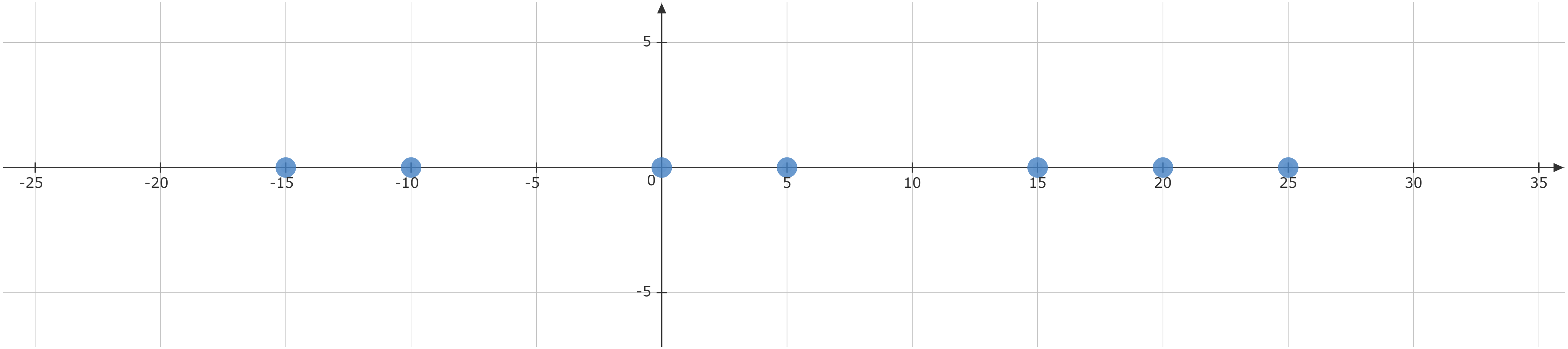
Companion Numericals

Consider the following one-dimensional dataset:

$$X = [-15,\ -10,\ 0,\ 5,\ 15,\ 20,\ 25]$$

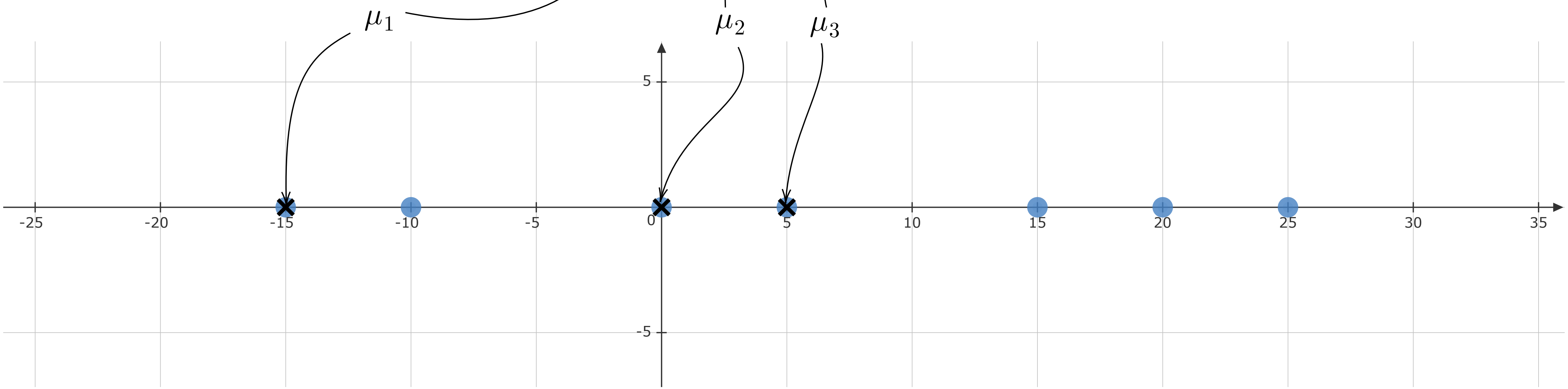
Consider the following one-dimensional dataset:

$$X = [-15, \ -10, \ 0, \ 5, \ 15, \ 20, \ 25]$$



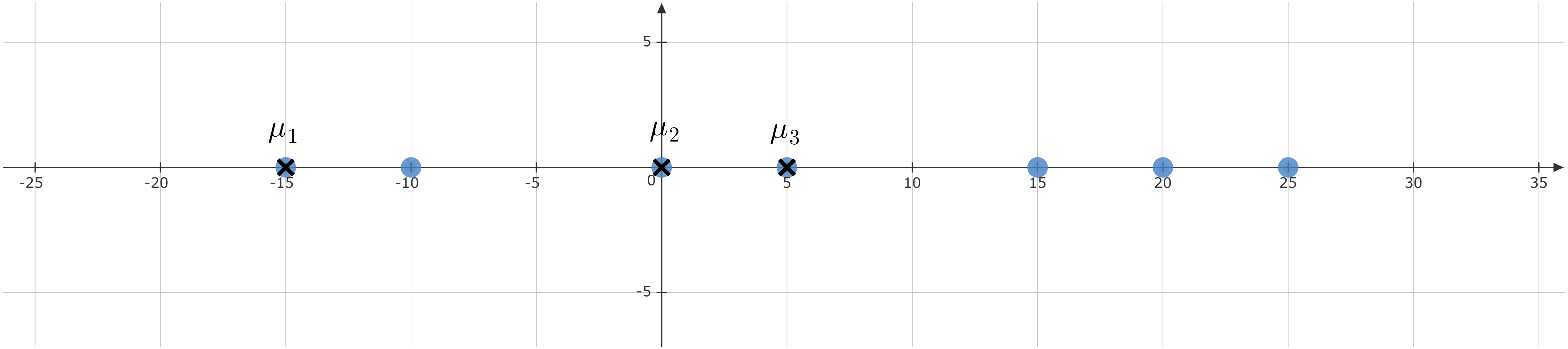
Consider the following one-dimensional dataset:

$$X = [-15,\ -10,\ 0,\ 5,\ 15,\ 20,\ 25]$$

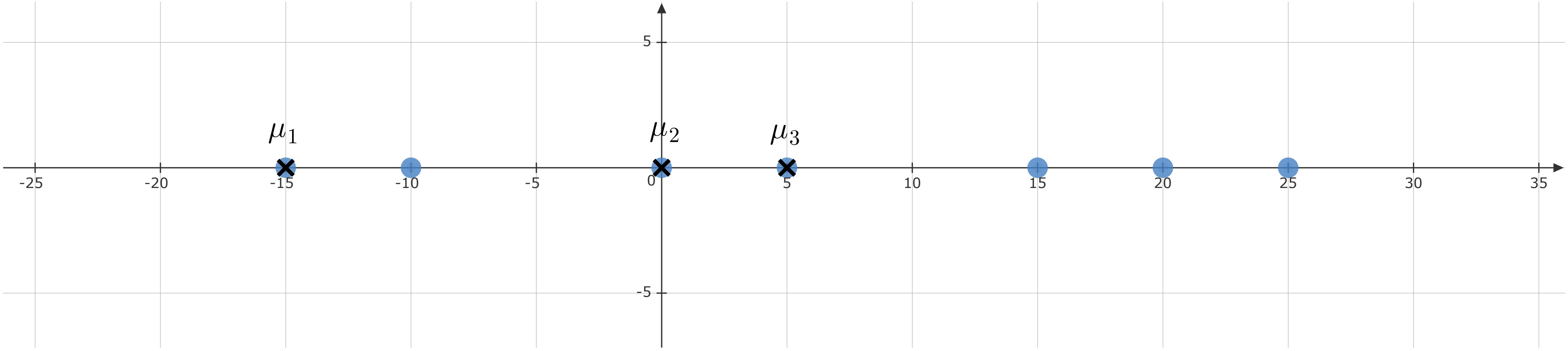


K-means algorithm with k=3 for the above cluster assignment is run on the given data points.

Cluster Assignment:



Cluster Assignment:

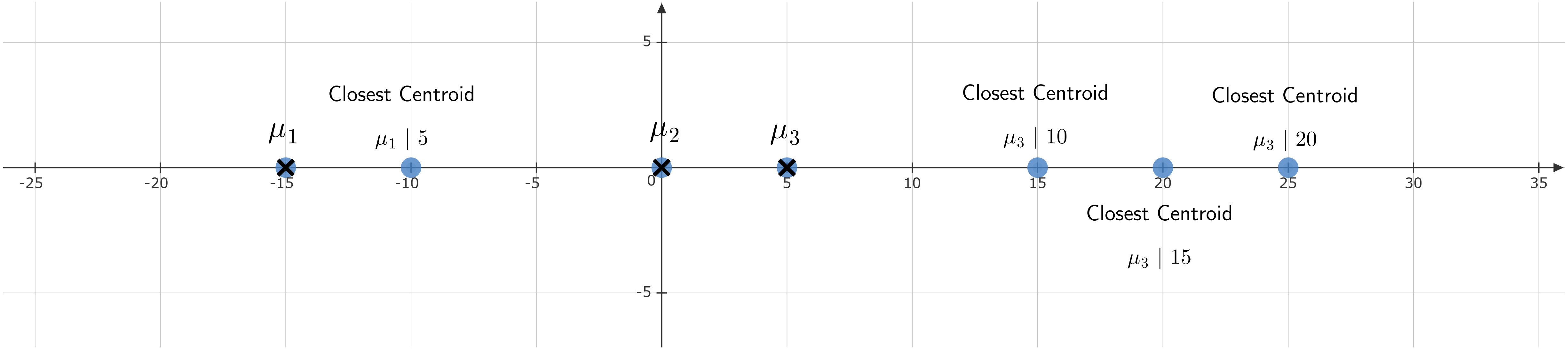


We need to find, for each point, the centre closest to the point.

Distance of $(-15,\ 0)$ from:

- $\mu_1 \ = \ \sqrt{(-15\ -\ (-15))^2\ +\ (0\ -\ 0)^2}\ =\ 5$
- $\mu_2 \ = \ \sqrt{(-15\ -\ (0))^2\ +\ (0\ -\ 0)^2}\ =\ 15$
- $\mu_3 \ = \ \sqrt{(-15\ -\ (5))^2\ +\ (0\ -\ 0)^2}\ =\ 20$

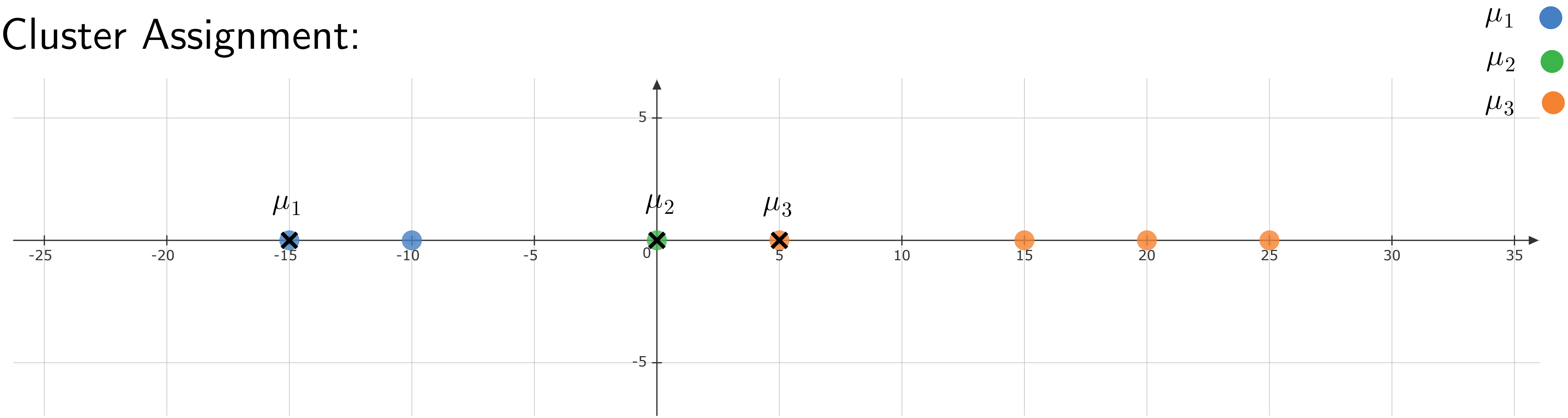
Cluster Assignment:



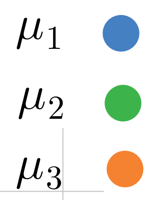
Center\Point	-15	-10	0	5	15	20	25
$\mu_1 = -15$	0	5	15	20	30	35	40
$\mu_2 = 0$	15	10	0	5	15	20	25
$\mu_3 = 5$	20	15	5	0	10	15	20

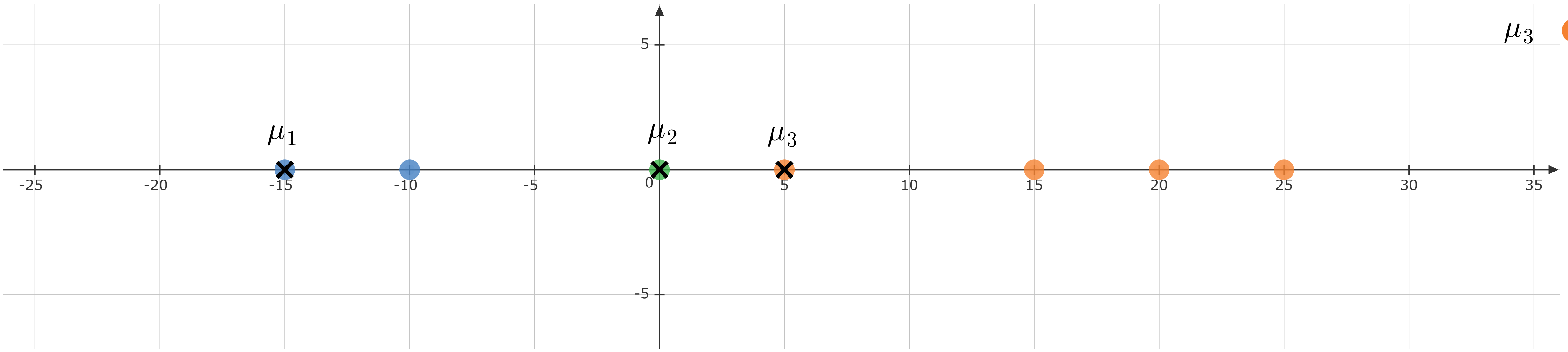
Iteration 1: Distance between points and cluster centers

Cluster Assignment:



Cluster Assignment:





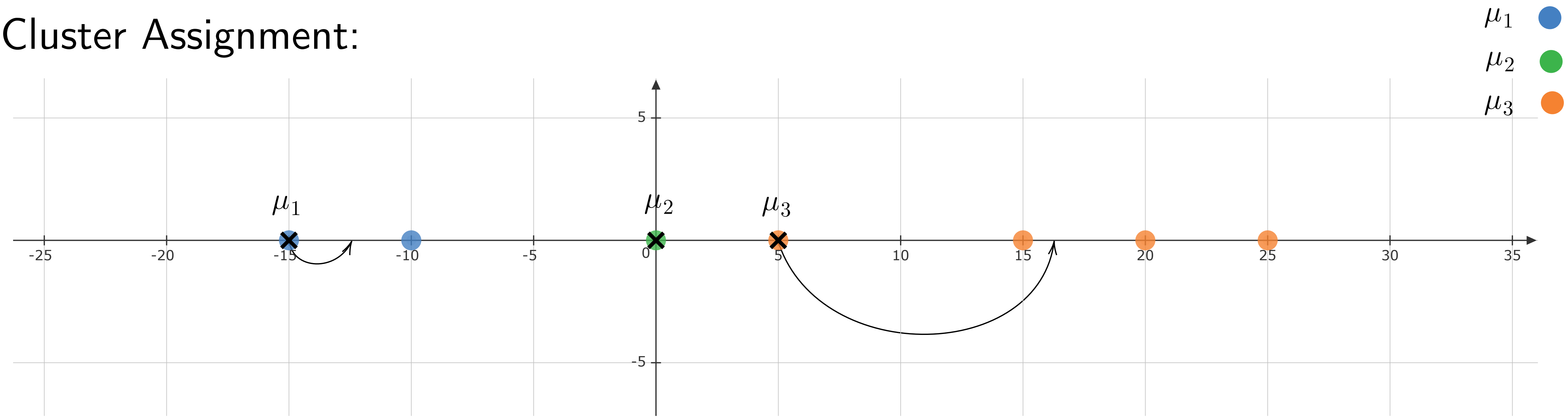
Cluster Computation:

$$\mu_1 = \frac{-15 - 10}{2} = -12.5$$

$$\mu_2 = \frac{0}{1} = 0$$

$$\mu_3 = \frac{5 + 15 + 20 + 25}{4} = 16.25$$

Cluster Assignment:



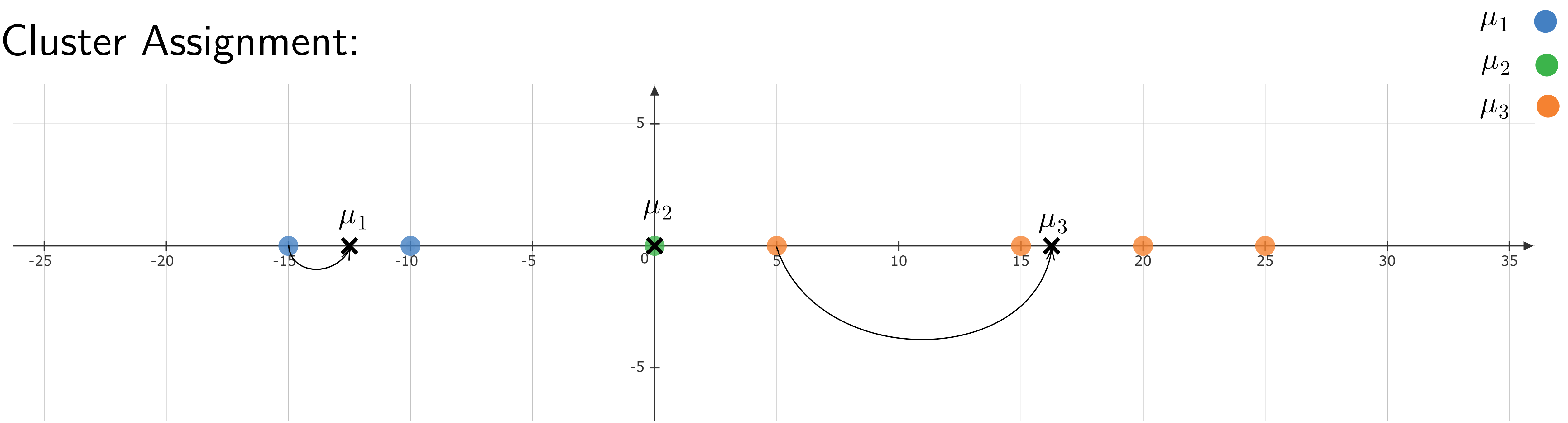
Cluster Computation:

$$\mu_1 = \frac{-15-10}{2} = -12.5$$

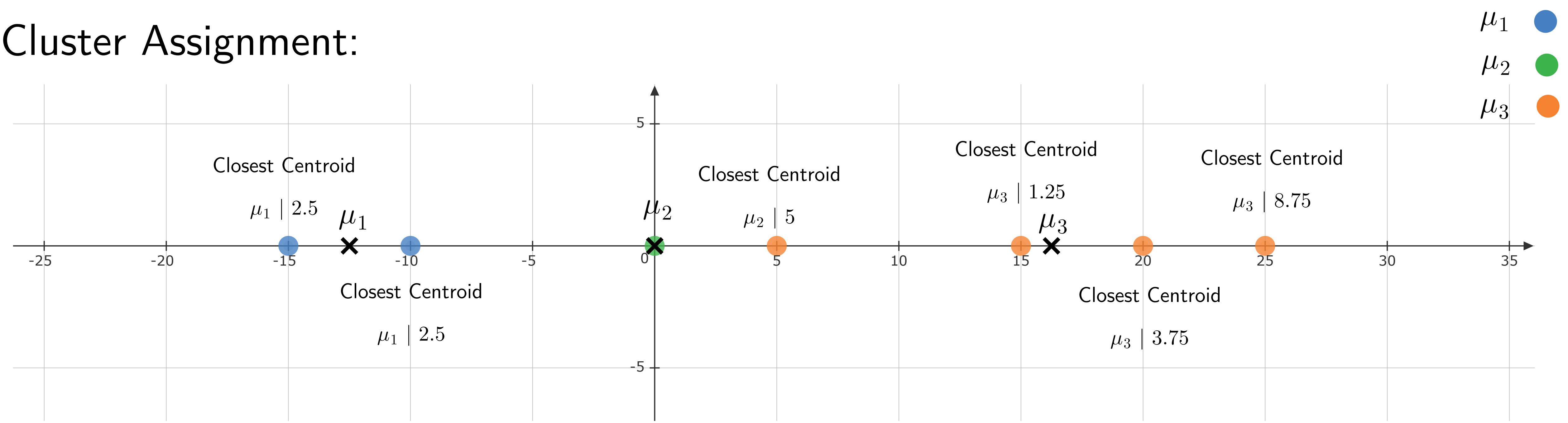
$$\mu_2 = \frac{0}{1} = 0$$

$$\mu_3 = \frac{5 + 15 + 20 + 25}{4} = 16.25$$

Cluster Assignment:



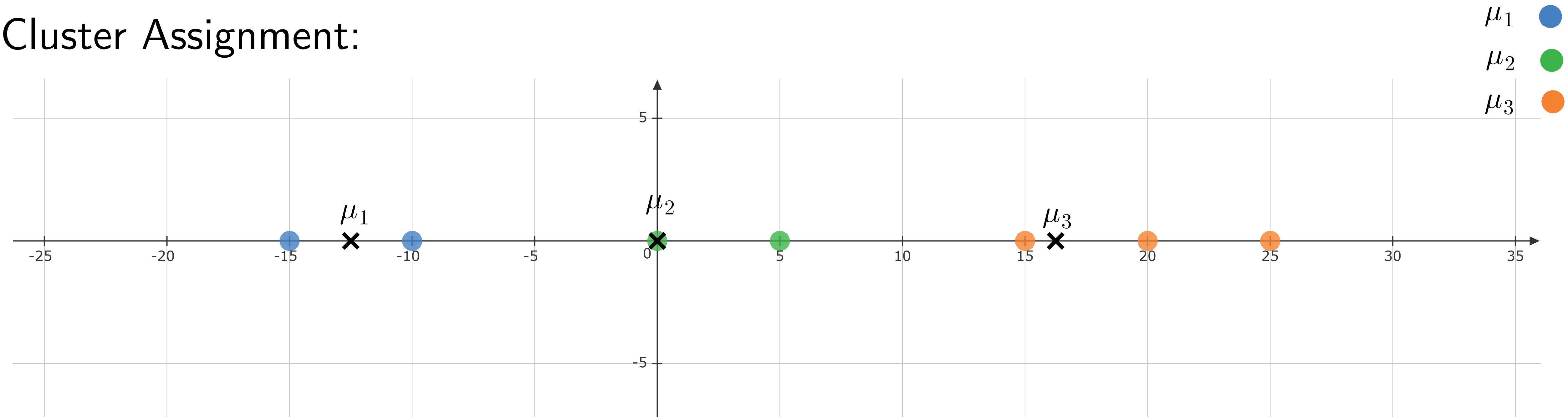
Cluster Assignment:



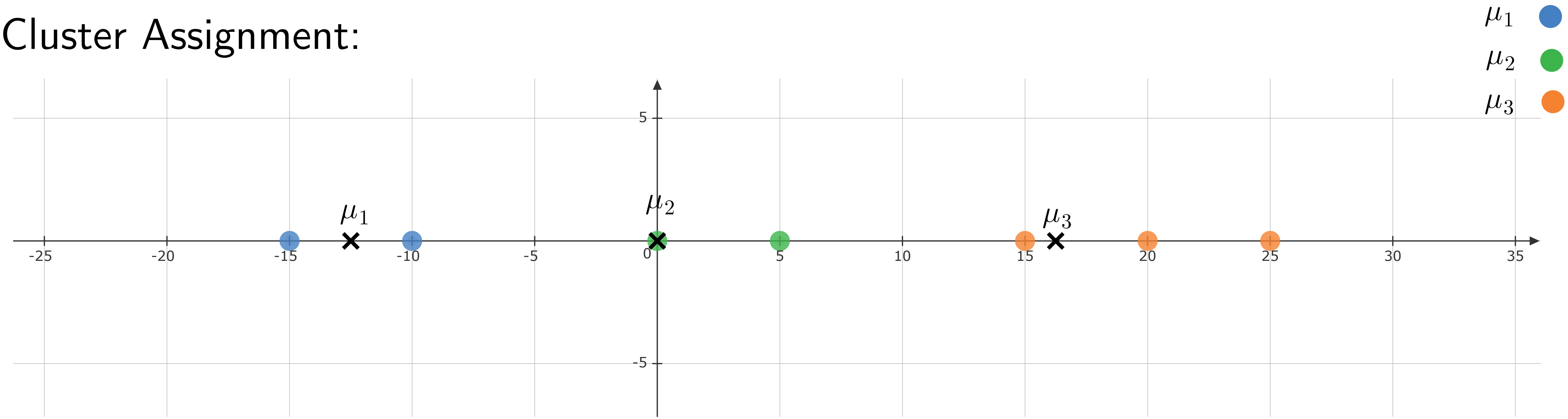
Center\Point	-15	-10	0	5	15	20	25
$\mu_1 = -12.5$	2.5	2.5	12.5	17.5	27.5	32.5	37.5
$\mu_2 = 0$	15	10	0	5	15	20	25
$\mu_3 = 16.25$	31.25	26.25	16.25	11.25	1.25	3.75	8.75

Iteration 2: Distance between points and cluster centers

Cluster Assignment:



Cluster Assignment:



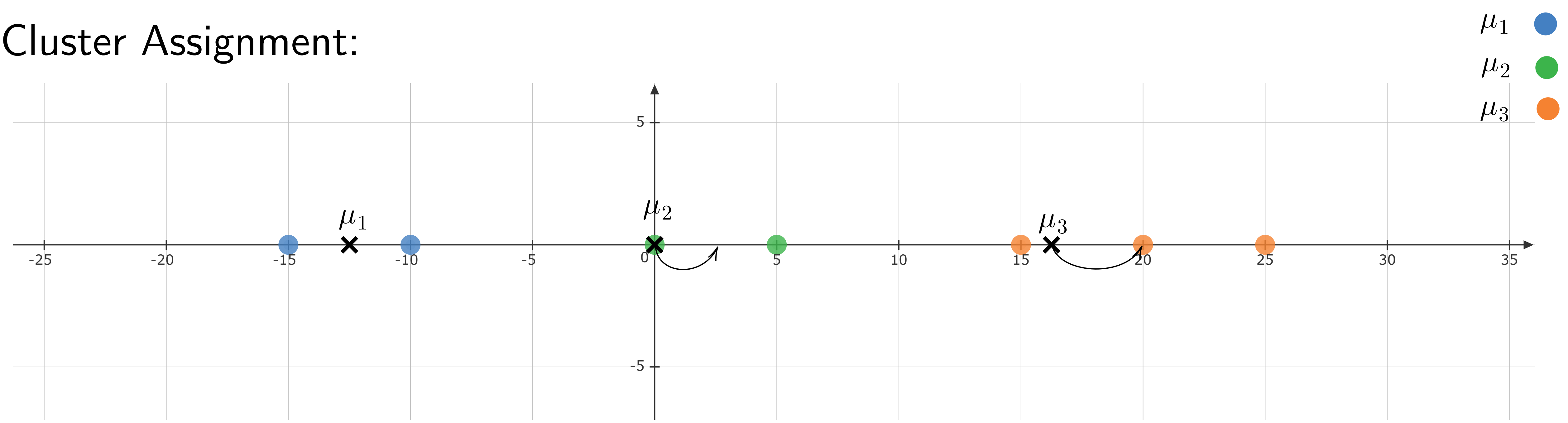
Cluster Computation:

$$\mu_1 = \frac{-15 - 10}{2} = -12.5$$

$$\mu_2 = \frac{0 + 5}{2} = 2.5$$

$$\mu_3 = \frac{15 + 20 + 25}{3} = 20$$

Cluster Assignment:



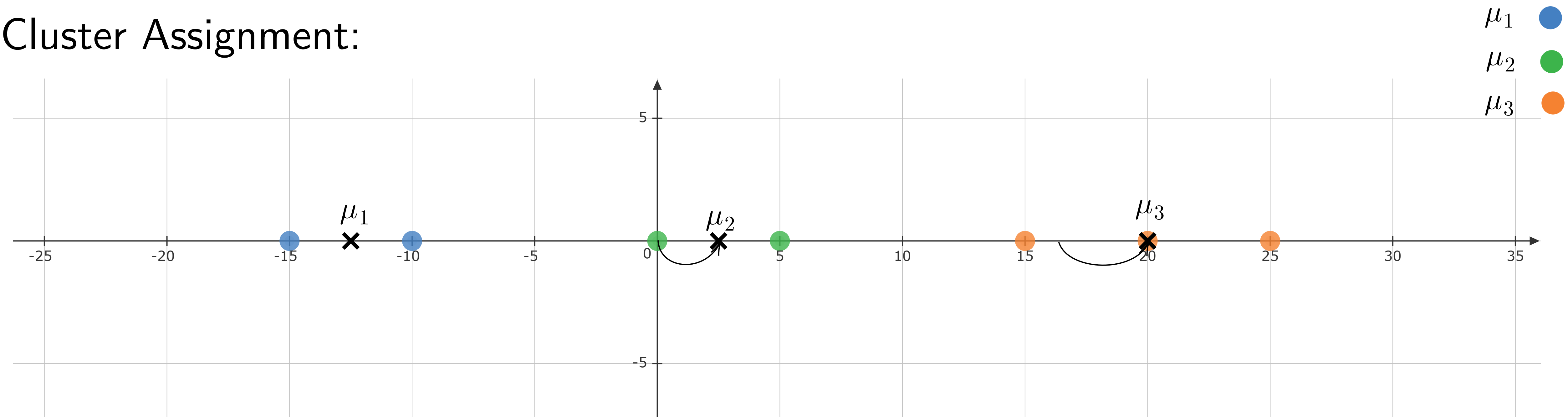
Cluster Computation:

$$\mu_1 = \frac{-15 - 10}{2} = -12.5$$

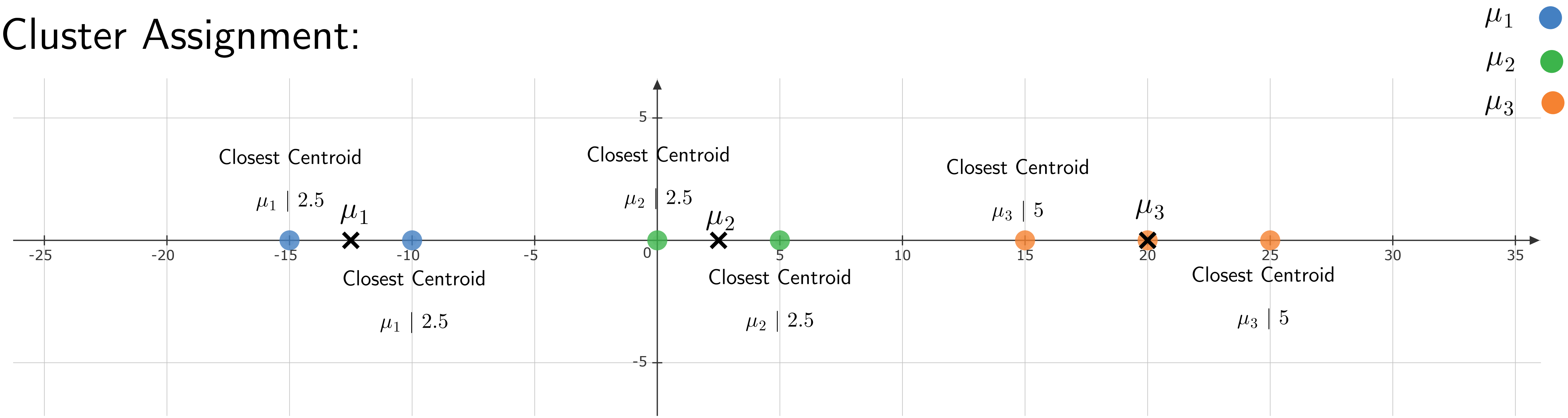
$$\mu_2 = \frac{0 + 5}{2} = 2.5$$

$$\mu_3 = \frac{15 + 20 + 25}{3} = 20$$

Cluster Assignment:



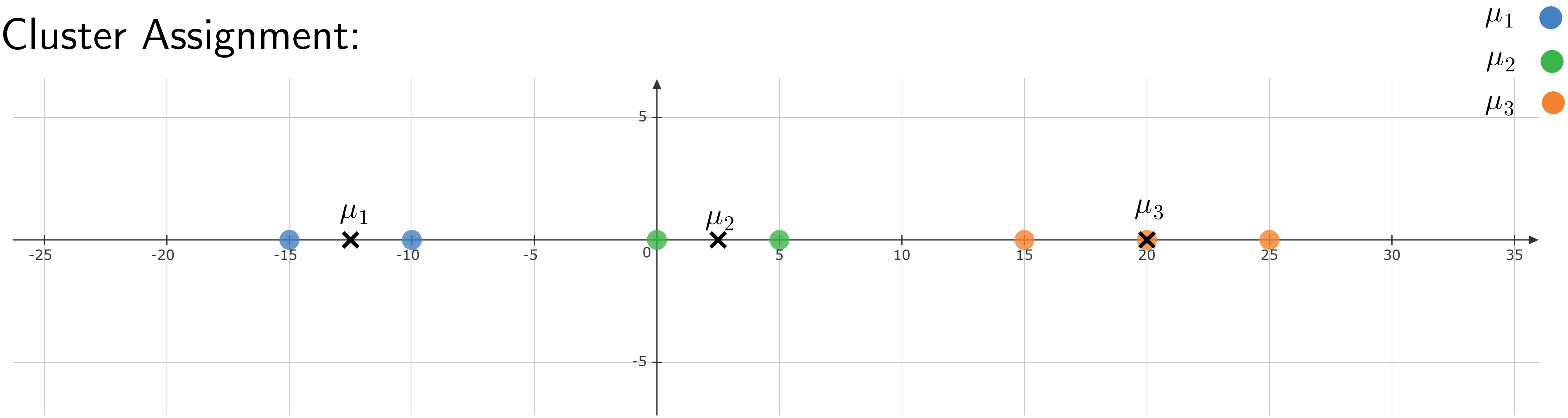
Cluster Assignment:



Center\Point	-15	-10	0	5	15	20	25
$\mu_1 = -12.5$	2.5	2.5	12.5	17.5	27.5	32.5	37.5
$\mu_2 = 2.5$	17.5	12.5	2.5	2.5	12.5	17.5	22.5
$\mu_3 = 20$	35	30	20	15	5	0	5

Iteration 3: Distance between points and cluster centers

Cluster Assignment:



The cluster assignments $\boxed{\text{do not change}}$.

The algorithm has therefore converged.

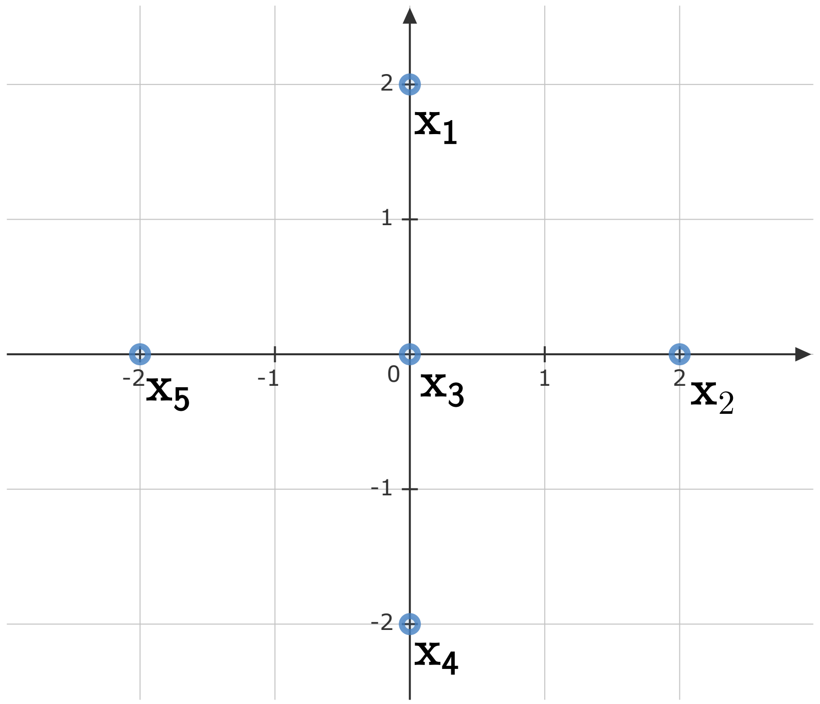
Consider the following dataset:

$$\left\{\mathbf{x}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \mathbf{x}_5 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}\right\}$$

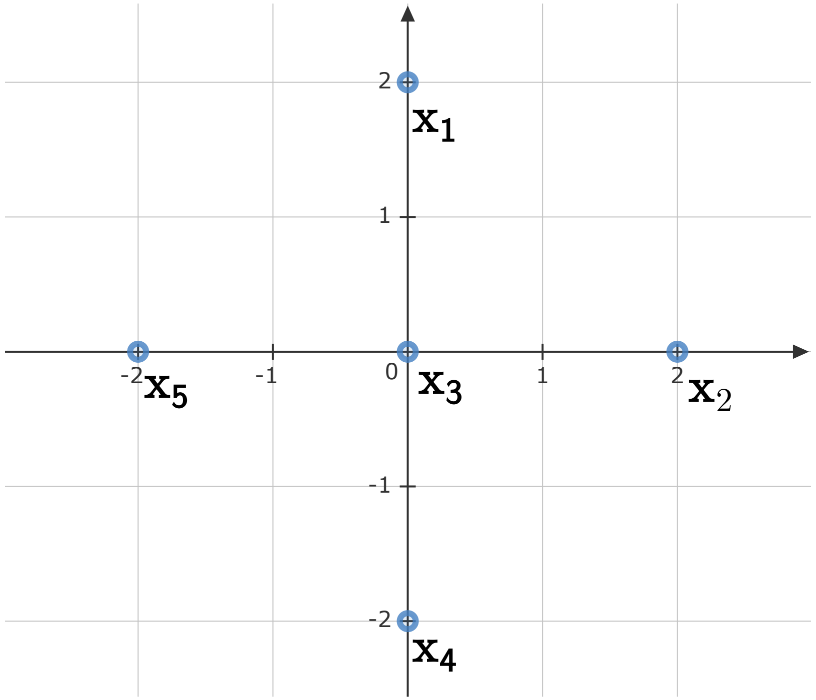
Consider the following dataset:

$$\left\{ \mathbf{x}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \mathbf{x}_5 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \right\}$$

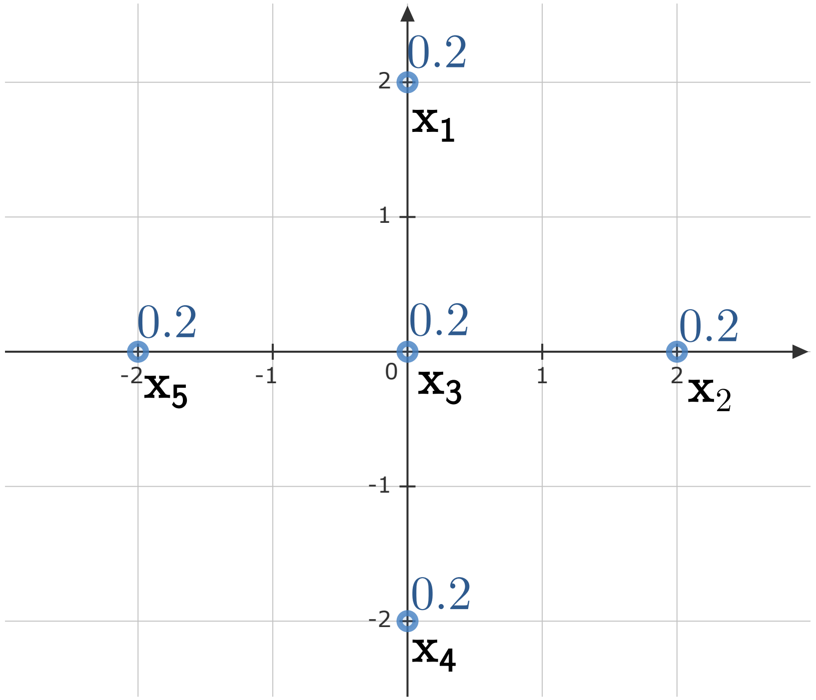
K-means++ Algorithm with k=2 is run on this dataset to initialize the cluster centers.



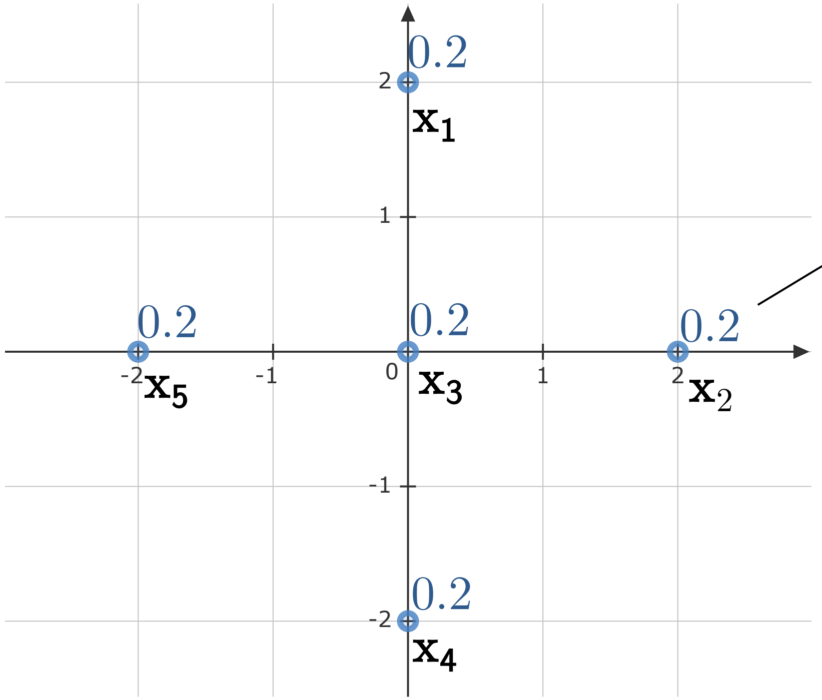
What is the probability of the following points $\mathbf{x}_2$, $\mathbf{x}_1$ (in that order) being chosen as initial cluster centers?



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



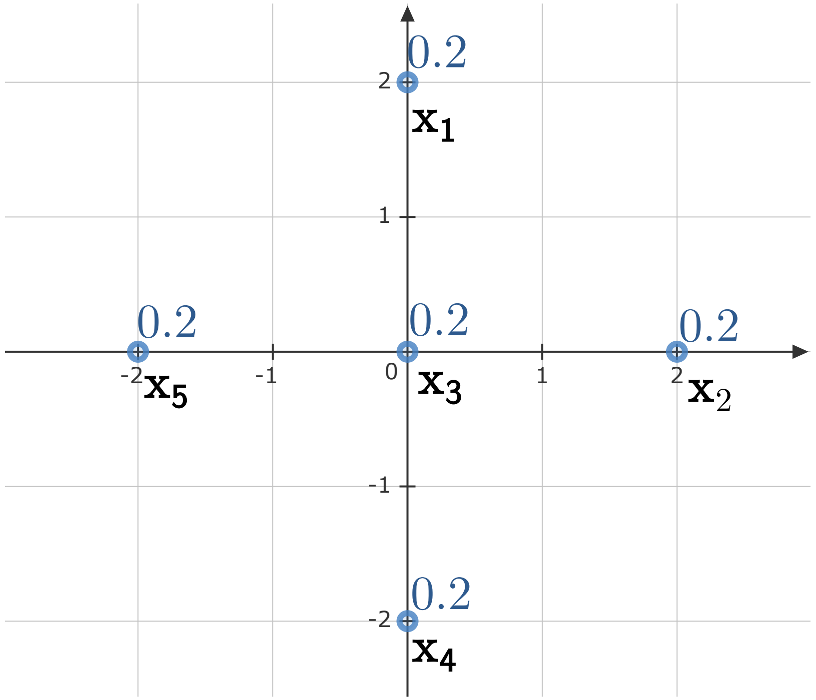
What is the probability of the following points $\mathbf{x}_2$, $\mathbf{x}_1$ (in that order) being chosen as initial cluster centers?



$$\mu = \begin{bmatrix} \quad \end{bmatrix}$$

$$P\left(\mu = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_1 \end{bmatrix}\right) =$$

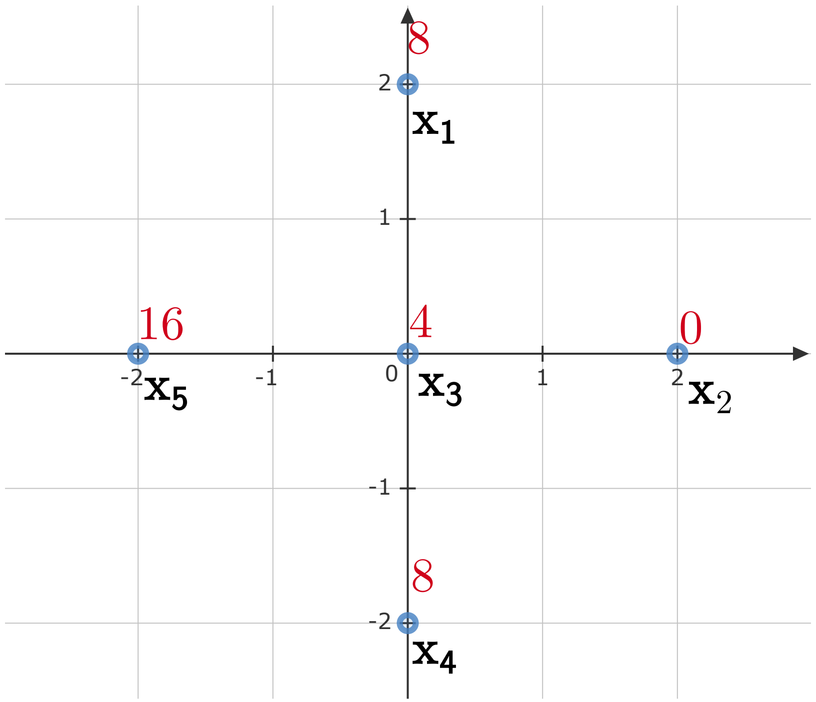
What is the probability of the following points $\mathbf{x}_2$, $\mathbf{x}_1$ (in that order) being chosen as initial cluster centers?



$$\mu = \begin{bmatrix} \mathbf{x}_2 & \end{bmatrix}$$

$$P\left(\mu = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_1 \end{bmatrix}\right) = 0.2 \times$$

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

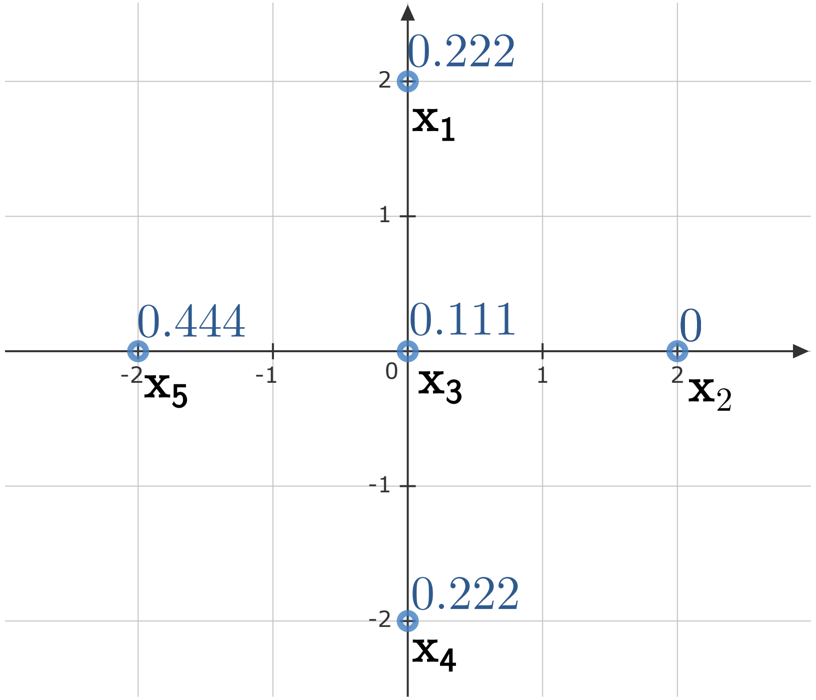


$$\mu = \left[\mathbf{x}_2 \quad \right]$$

$$P\left(\mu = \left[\mathbf{x}_2 \;\; \mathbf{x}_1\right]\right) = 0.2 \times$$

Compute score for each datapoint to be the minimum value of the set of squared distances between the datapoint and the chosen clusters.

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



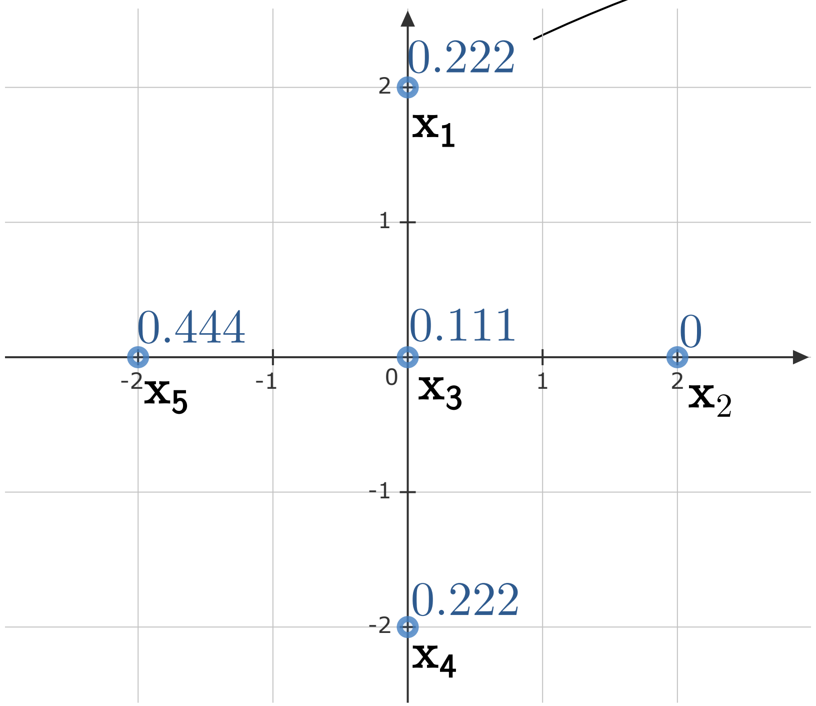
$$\mu = [\mathbf{x}_2 \quad]$$

$$P(\mu = [\mathbf{x}_2 \ \mathbf{x}_1]) = 0.2 \ \times$$

Normalize scores $P(\mathbf{x}_i) = \frac{S_i}{\sum S_i}$

Where S_i denotes the score for $\mathbf{x}_i$

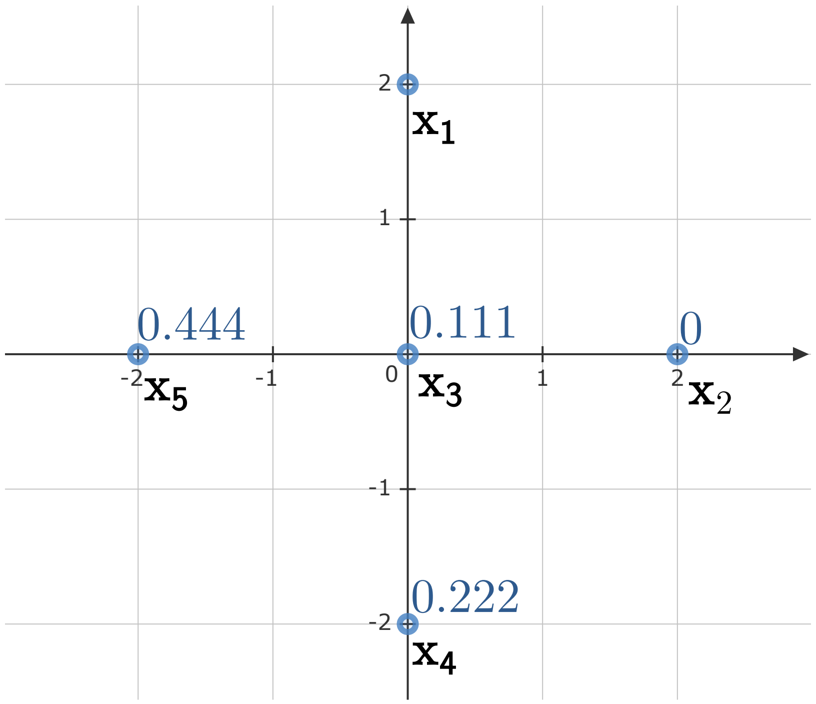
What is the probability of the following points $\mathbf{x}_2$, $\mathbf{x}_1$ (in that order) being chosen as initial cluster centers?



$$\mu = [\mathbf{x}_2 \quad]$$

$$P\big(\mu = [\mathbf{x}_2 \;\; \mathbf{x}_1]\big) = 0.2 \times$$

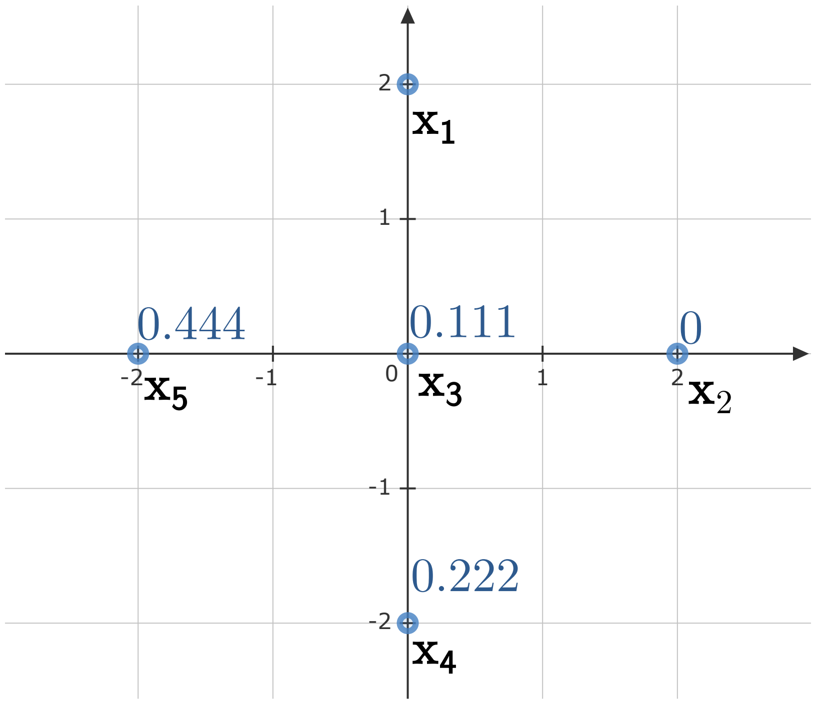
What is the probability of the following points $\mathbf{x}_2$, $\mathbf{x}_1$ (in that order) being chosen as initial cluster centers?



$$\mu = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_1 \end{bmatrix}$$

$$P\left(\mu = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_1 \end{bmatrix}\right) = 0.2 \times 0.222$$

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



$$\mu = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_1 \end{bmatrix}$$

$$P\left(\mu = \begin{bmatrix} \mathbf{x}_2 & \mathbf{x}_1 \end{bmatrix}\right) = 0.2 \times 0.222 \approx \boxed{0.044}$$