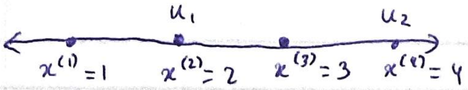
1. Data points in $\mathbb{R}$: $x^{(1)} = 1$, $x^{(2)} = 2$, $x^{(3)} = 3$, $x^{(4)} = 4$

$u_1 = 2$ and $u_2 = 4$

Assume that if point $x^{(i)}$ is equally distant to multiple centroids u_k, the point is assigned to the centroid whose index is smallest.



<u>k-means:</u>

① Cluster Assignment step: $c^1 = 1$, $c^2 = 1$, $c^3 = 1$, $c^4 = 2$

② Move Centroid: $u_1 = \frac{1+2+3}{3} = 2$

$u_2 = \frac{4}{1} = 4$

As cluster centroids are the same, the algorithm ends.

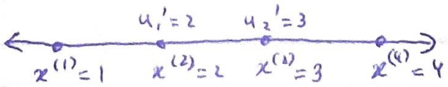
Distortion: $J = \frac{1}{m} \sum_{i=1}^{m} \| x^{(i)} - u_{c^{(i)}} \|^2 = \frac{1}{4} \left[(1-2)^2 + (2-2)^2 + (3-2)^2 + (4-4)^2 \right]$

$= \frac{1}{4} \left[1 + 0 + 1 + 0 \right]$

$= \frac{1}{2}$

Let's verify if $u_1 = 2$ and $u_2 = 4$ is the globally optimal solution by checking if we get a lower distortion at different cluster centroids.

$u_1' = 2$ and $u_2' = 3$



k-means:

① $c^1 = 1$, $c^2 = 1$, $c^3 = 2$, $c^4 = 2$

② $u_1' = \frac{1+2}{2} = \frac{3}{2}$ $\quad u_2' = \frac{3+4}{2} = \frac{7}{2}$

Repeat, as centroids are different

① $c^1 = 1$, $c^2 = 1$, $c^3 = 2$, $c^4 = 2$

② $u_1'' = \frac{1+2}{2} = \frac{3}{2}$, $u_2'' = \frac{3+4}{2} = \frac{7}{2}$

Stop algorithm, as $u_1'' = u_1'$ and $u_2'' = u_2'$.

Find distortion J'.

$$J' = \frac{1}{m} \sum_{i=1}^{m} \| x^{(i)} - u_{c^{(i)}} \|^2$$

$$J' = \frac{1}{4} \left[\left(1 - \frac{3}{2}\right)^2 + \left(2 - \frac{3}{2}\right)^2 + \left(3 - \frac{7}{2}\right)^2 + \left(4 - \frac{7}{2}\right)^2 \right]$$

$$J' = \frac{1}{4} \left[(-0.5)^2 + (0.5)^2 + (-0.5)^2 + (0.5)^2 \right]$$

$$J' = \frac{1}{4}$$

Thus $J = \frac{1}{2} > J' = \frac{1}{4}$ which proves the cluster assignment $u_1' = \frac{3}{2}$, $u_2' = \frac{7}{2}$ is more optimal than $u_1 = 2$, $u_2 = 4$.