Part-2 Lack of Optimality of k-means

$x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$

$\mu_1 = 2$ & $\mu_2 = 4$

number of cluster $k = 2$

$X = \{1, 2, 3, 4\}$

$\mu_1 = 2$, $\mu_2 = 4$

$k = 2$

Assign cluster based on distance.

$\mu_1 = 2$	$\mu_2 = 4$
for $x_1 = 1 - 2 = 1$	for $x_1 \Rightarrow 1 - 4 = 3$
for $x_2 = 2 - 2 = 0$	for $x_2 \Rightarrow 2 - 4 = 2$
for $x_3 = 3 - 2 = 1$	for $x_3 \Rightarrow 3 - 4 = 1$
for $x_4 = 4 - 2 = 2$	for $x_4 \Rightarrow 4 - 4 = 0$.

so here x_1, x_2 directly assign to cluster 1 with centroid μ_1 & x_4 is directly assign to cluster 2 with centroid μ_2

For x_3 distance from both centroids is same but as per Note we will take x_3 to μ_1 because of instruction give in Note.

cluster I = {1, 2, 3}
cluster 2 = {4}.

Update Centroid.

$\mu_1^{new} \Rightarrow \frac{1+2+3}{3} = 2 = \mu_1$

$\mu_2^{new} \Rightarrow \frac{4}{1} = 4 = \mu_2.$

Here $\mu_1^{new} = \mu_1$, $\mu_2^{new} = \mu_2$ $\Rightarrow$ centroids don't change.

Convergence check or check for gobally optimal.

$$\text{distortion} = \sum_{x_i \in C_1} (x_i - \mu_K)^2 + \sum_{x_i \in C_2} (x_i - \mu_K)^2$$

$$= (1-2)^2 + (2-2)^2 + (3-2)^2 + (4-4)^2.$$

$$\Rightarrow 1 + 0 \; 1 + 0$$

$$= 2$$

Compare with global optimal.

let's take.

Cluster I = {1, 2} with $\mu_1 = 2$
Cluster 2 = {3, 4} with $\mu_2 = 4$.

Centroid Update.

$$\mu_1^{new} = \frac{1+2}{2} = 1.5 < \mu_1$$

$$\mu_2^{new} \Rightarrow \frac{4+3}{2} = 3.5 < \mu_2.$$

Here $\boxed{\begin{array}{l} \mu_1^{new} < \mu_1 \\ \mu_2^{new} < \mu_2 \end{array}}$

distortion :-

As per previous formula.

$$(1-1.5)^2 + (2-1.5)^2 + (3-3.5)^2 + (4-3.5)^2$$

$$\Rightarrow 0.25 + 0.25 + 0.25 + 0.25$$

$$\Rightarrow 1 < 2.$$

So Here we found distortion $I < 2$

$\Rightarrow$ To conclude this show k-means might converge to a solution that is not globally optimal.