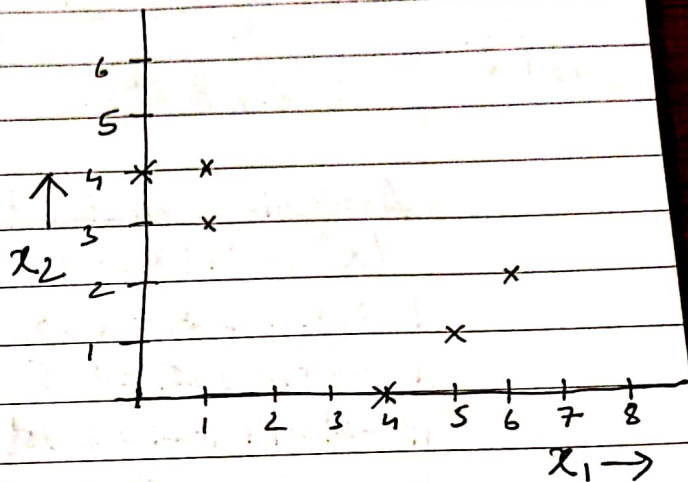


Problem 4: K-Means Clustering

i	$x_1^{(i)}$	$x_2^{(i)}$	k	k_1
1	1	4	1	1
2	1	3	2	1
3	0	4	1	1
4	5	1	1	2
5	6	2	1	1
6	4	0	2	2



Calculating Centroid for Clusters

Centroid for C_1 ($k=1$)

$$\left(\frac{1+0+5+6}{4}, \frac{4+4+1+2}{4} \right)$$

$$\left(\frac{12}{4}, \frac{11}{4} \right)$$

$$(3, 2.75)$$

Centroid for C_2 ($k=2$)

$$\left(\frac{1+4}{2}, \frac{3+0}{2} \right)$$

$$(2.5, 1.5)$$

Calculating distance using Manhattan Method

$ x_1 - 3 + x_2 - 2.75 $	dist	$ x_1 - 2.5 + x_2 - 1.5 $
$ 1-3 + 4-2.75 = 2 + 1.25 = 3.25$	<	$ 1-2.5 + 4-1.5 = 4$
$ 1-3 + 3-2.75 = 2 + 0.25 = 2.25$	<	$ 1-2.5 + 3-1.5 = 3$
$ 0-3 + 4-2.75 = 3 + 1.25 = 4.25$	<	$ 0-2.5 + 4-1.5 = 5$
$ 5-3 + 1-2.75 = 3.75$	>	$ 5-2.5 + 1-1.5 = 3$
$ 6-3 + 2-2.75 = 3.75$	<	$ 6-2.5 + 2-1.5 = 4$
$ 4-3 + 0-2.75 = 3.75$	>	$ 4-2.5 + 0-1.5 = 3$

Assigning datapoints, based on the distance, to the cluster. (k_1)

Centroid for Cluster 1 (updated) → Calculating new centroids based on new clusters

$$\left(\frac{1+1+0+6}{4}, \frac{4+3+4+2}{4} \right) = (2, 3.25)$$

Centroid for Cluster 2 (updated) →

$$\left(\frac{5+4}{2}, \frac{1+0}{2} \right) = (4.5, 0.5)$$

Assigning cluster data points to clusters based on new centroids

$ x_1 - 2 + x_2 - 3.25 $		$ x_1 - 4.5 + x_2 - 0.5 $	updated cluster
$ 1-2 + 4-3.25 = 1.75$	<	$ 1-4.5 + 4-0.5 = 3.5$	1
$ 1-2 + 3-3.25 = 1.25$	<	$ 1-4.5 + 3-0.5 = 2.5$	1
$ 0-2 + 4-3.25 = 2.75$	<	$ 0-4.5 + 4-0.5 = 3.5$	1
$ 5-2 + 1-3.25 = 5.25$	>	$ 5-4.5 + 1-0.5 = 0.5$	2
$ 6-2 + 2-3.25 = 5.25$	>	$ 6-4.5 + 2-0.5 = 1.5$	2
$ 4-2 + 0-3.25 = 5.25$	>	$ 4-4.5 + 0-0.5 = 1$	2

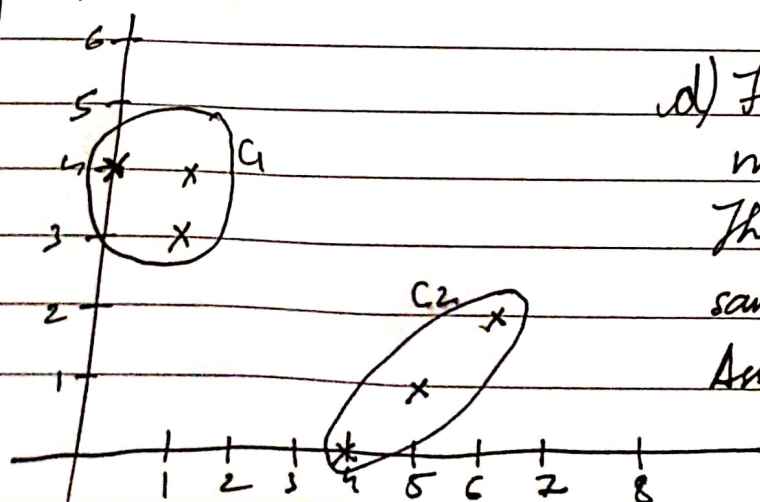
Centroid for C_1

$$C_1 = \left(\frac{1+1+0}{3}, \frac{4+3+4}{3} \right) = \left(\frac{2}{3}, \frac{11}{3} \right) = (0.66, 3.66)$$

$$C_2 = \left(\frac{5+6+4}{3}, \frac{1+2+0}{3} \right) = (5, 1)$$

$ x_1 - 0.66 + x_2 - 3.66 $		$ x_1 - 5 + x_2 - 1 $	updated cluster (k3)
$ 1-0.66 + 4-3.66 = 0.66$	<	$ 1-5 + 4-1 = 7$	1
$ 1-0.66 + 3-3.66 = 0.99$	<	$ 1-5 + 3-1 = 6$	1
$ 0-0.66 + 4-3.66 = 0.99$	<	$ 0-5 + 4-1 = 8$	1
$ 5-0.66 + 1-3.66 = 6.99$	>	$ 5-5 + 1-1 = 0$	2
$ 6-0.66 + 2-3.66 = 6.99$	>	$ 6-5 + 2-1 = 2$	2
$ 4-0.66 + 0-3.66 = 6.99$	>	$ 4-5 + 0-1 = 2$	2

As the centroids & the cluster labels stop changing, updated clusters are - (k3)



d) Final clustering does not match with initial clustering. The ~~cluster~~ clustering result always same regardless of initial cluster Assignment as shown in the R-Code.