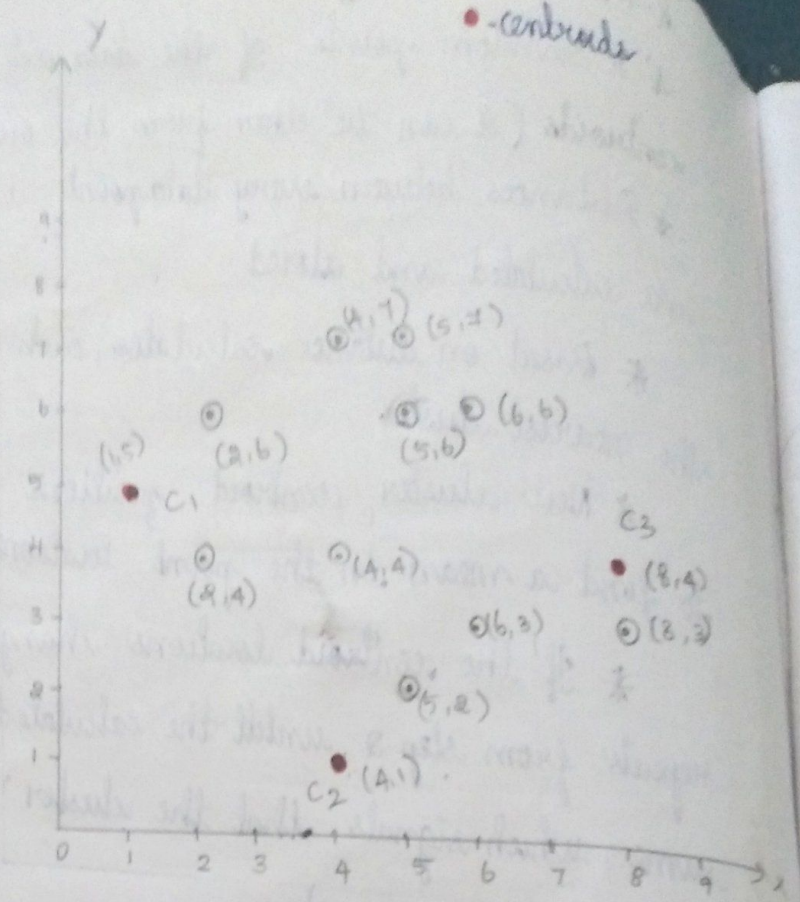
K-Means Clustering:

* K random points of the data set are chosen to be centroids (It can be other from the output dataset).
* Distances between every data point and the k centroids are calculated and stored.
* Based on distance calculated, each point is assigned to the nearest cluster
* New cluster centroid positions are updated: similar to find a mean in the point locations.
* If the centroid locations changed, the process repeats from step 2, until the calculated new centre stays the same, which signals that the cluster's members and centroids are now set.

Clustering exercise:

X_1	X_2
2	4
2	6
5	6
4	7
8	3
6	6
5	2
5	7
6	3
4	4



K = 3

Iteration 1:

C1 - seed point 1 - (1,5)

C2 - seed point 2 - (4,1)

C3 - seed point 3 - (8,4)

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

X_1	X_2	Distance to (1,5)	Distance to (4,1)	Distance to (8,4)	cluster number
2	4	1.41	3.61	6.00	C1
2	6	1.41	5.39	6.32	C1
5	6	4.12	5.10	3.61	C3
4	7	3.61	6.00	5.00	C1
8	3	7.28	4.47	1.00	C3
6	6	5.10	5.39	2.83	C3
5	2	5.00	1.41	3.61	C2
5	7	4.47	6.08	4.24	C3
6	3	5.39	2.83	2.24	C3
4	4	3.16	3.00	4.00	C2

$$C1\text{-centroid} = \left(\frac{2+2+4}{3}, \frac{4+6+7}{3}\right)$$

$$C1 = (2.66, 5.66)$$

$$C2\text{-centroid} = \left(\frac{5+4}{2}, \frac{2+4}{2}\right)$$

$$C2 = (4.5, 3)$$

$$C3\text{-centroid} = \left(\frac{5+8+6+5+6}{5}, \frac{6+3+6+7+3}{5}\right)$$

$$C3 = (6, 5)$$

Iteration - 2:

C1 - centroid - (2.66, 5.66)

C2 - centroid - (4.5, 3)

C3 - centroid - (6, 5).

X	Y	distance to (2.66, 5.66)	distance to (4.5, 3)	distance to (6, 5)	Cluster Number
2	4	1.79	2.69	4.12	C1
2	6	0.74	3.91	4.12	C1
5	6	2.36	3.04	1.41	C3
4	7	1.90	4.03	2.83	C1
8	3	5.97	3.5	2.83	C3
6	6	3.36	3.35	1	C3
5	2	4.34	1.12	3.16	C2
5	7	2.70	4.03	2.24	C3
6	3	4.27	1.5	2	C2
4	4	2.13	1.118	2.24	C2

C1-centroid = $\left(\frac{2+2+4}{3}, \frac{4+6+7}{3}\right) = (2.66, 5.66)$

C2-centroid = $\left(\frac{6+4+5}{3}, \frac{3+4+2}{3}\right) = (5, 3)$

C3-centroid = $\left(\frac{5+8+6+5}{4}, \frac{6+3+6+7}{4}\right) = (6, 5.5)$

Iteration - 3 :

X	Y	Distance to (2.66, 5.66)	Distance to (5, 3)	Distance to (6, 5.5)	cluster Number
2	4	1.79	3.16	4.27	C1
2	6	0.74	4.84	4.03	C1
5	6	2.36	3.00	1.12	C3
4	7	1.90	4.12	2.50	C1
8	3	5.97	3.00	3.20	C2
6	6	3.36	3.16	0.50	C3
5	2	4.34	1.00	3.64	C2
5	7	2.70	4.00	1.80	C3
6	3	4.27	1.00	2.50	C2
4	4	2.13	1.41	2.5	C2
4	4				

C1 - centroid = (2.66, 5.66)

C2 - centroid = (5.75, 3)

C3 - centroid = (5.33, 6.33)

Iteration 4:

C1 - centroid - (2.66, 5.66)

C2 - centroid - (5.75, 3)

C3 - centroid - (5.33, 6.33)

X	Y	Distance to (2.66, 5.66)	Distance to (5.75, 3)	Distance to (5.33, 6.33)	cluster number
2	4	1.79	3.88	4.06	C1
2	6	0.74	4.80	3.35	C1
5	6	2.36	3.09	0.47	C3
4	7	1.90	4.37	1.49	C3
8	3	5.97	2.25	4.27	C2
6	6	3.36	3.01	0.75	C3
5	2	4.34	1.25	4.34	C2
5	7	2.70	4.07	0.75	C3
6	3	4.27	0.25	3.40	C2
4	4	2.13	2.015	2.68	C2

C1 - centroid - (2, 5)

C2 - centroid - (5.75, 3)

C3 - centroid - (5, 6.5).

Iteration 5:

C1 - centroid - (2, 5)

C2 - centroid - (5.75, 3)

C3 - centroid - (5, 6.5)

X	Y	Distance to (2,5)	Distance to (5.75, 3)	Distance to (5, 6.5)	cluster number
2	4	1.00	3.88	3.91	C1
2	6	1.00	4.80	3.04	C1
5	6	3.16	3.09	0.50	C3
4	7	2.83	4.37	1.12	C3
8	3	6.32	2.25	4.61	C2
6	6	4.12	3.01	1.12	C3
5	2	4.24	1.25	4.50	C2
5	7	3.61	4.07	0.50	C3
6	3	4.47	0.25	3.64	C2
4	4	2.24	2.02	2.69	C2

No movements of data points.

Hence these are the final positions.

(6, 2)

(5, 6)

(2, 4)