1. DATASET:

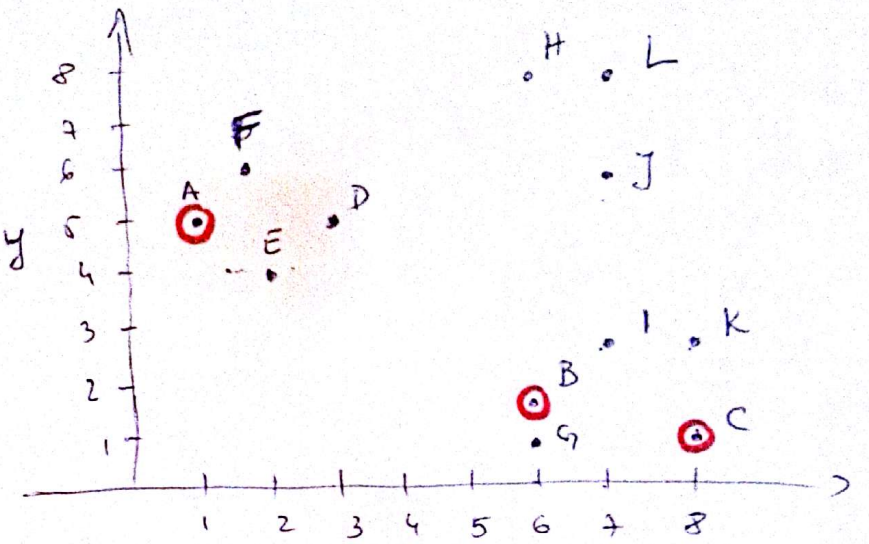
X	y	
1	5	A
6	2	B
8	1	C
3	5	D
2	4	E
2	6	F
6	1	G
6	8	H
7	3	I
7	6	J
8	3	K
8	7	L

CLUSTERING

metric – L_2 (Euclidean distance)

$K = 3$

USE FIRS three datapoints as centroids



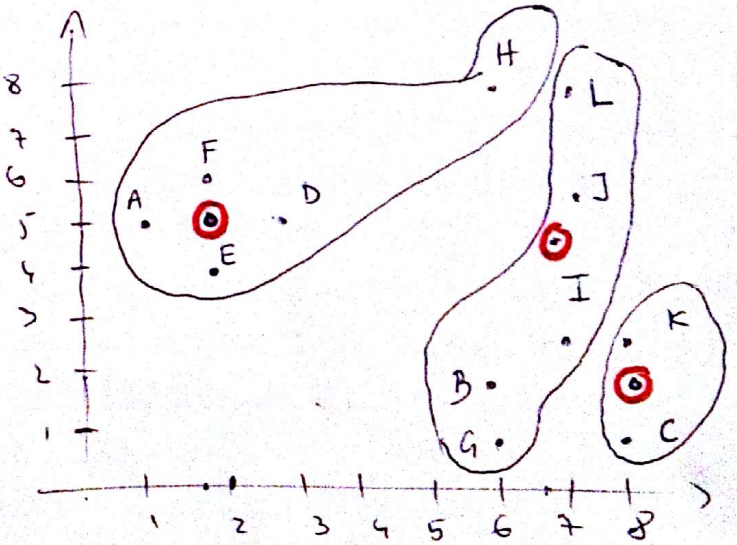
	L	A CLUSTER 1	B CLUSTER 2	C CLUSTER 3	CLUSTER ↓
0	A	0	5,83	8,06	1
1	B	5,83	0	2,23	2
2	C	8,06	2,23	0	3
3	D	2	4,24	6,4	1
4	E	1,41	4,47	6,7	1
5	F	1,41	5,65	7,81	1
6	G	6,4	1	2	2
7	H	5,83	6	7,28	1
8	I	6,32	1,41	2,23	2
9	J	6,08	4,12	5,09	2
10	K	7,28	2,23	2	3
11	L	7,28	5,38	6	2

$d(A,B) = \sqrt{(1-6)^2 + (5-2)^2} = \sqrt{34} = 5,83$

$d(A,C) = \sqrt{(1-8)^2 + (5-1)^2} = \sqrt{65} = 8,06$

$d(B,A) = \sqrt{\dots} = 5,83$

ON THE REPO, A FUNCTION FOR CALCULATING DISTANCES ~~HAS~~ CAN BE FOUND AS A PYTHON NOTEBOOK



CLUSTER 1 – $\{A, D, E, F, H\}$

CLUSTER 2 – $\{B, G, I, J, L\}$

CLUSTER 3 – $\{C, K\}$

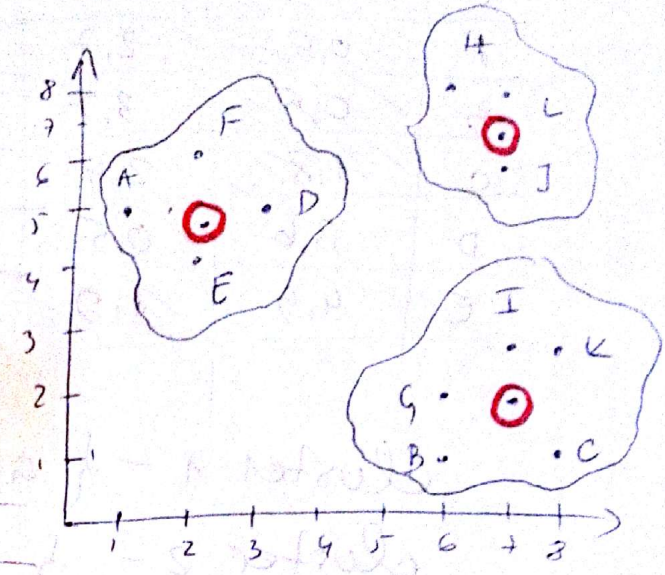
$\bar{x}_1 = (1,6; 5)$

$\bar{x}_2 = (6,66; 4,5)$

$\bar{x}_3 = \left(\frac{8+8}{2}, \frac{1+3}{2}\right) = (8,2)$

Scanned by CamScanner

i		X_1^1	X_2^1	X_3^1	CLUSTER
0	A	~~4,66~~	5,9	7,61	1
1	B	~~5,22~~	4,71	~~2~~	3
2	C	~~7,29~~	5,82	~~1~~	3
3	D	~~1,22~~	4,02	5,83	1
4	E	~~1,05~~	5,37	6,32	1
5	F	~~1,66~~	4,71	7,21	1
6	G	~~5,85~~	5,7	~~2,23~~	3
7	H	~~5,22~~	~~1,49~~	6,32	2
8	I	~~5,38~~	3,68	~~1,41~~	3
9	J	~~5,09~~	~~0,74~~	4,12	2
10	K	~~6,32~~	3,9	~~1~~	3
11	L	~~6,32~~	~~1,37~~	5	2



CLUSTER 1 - {A, D, E, F} $\quad \bar{X}_1^2 - (2, 4.5)$

CLUSTER 2 - {H, J, L} $\quad \bar{X}_2^2 - (7, 7)$

CLUSTER 3 - {B, G, I, K, C} $\quad \bar{X}_3^2 - (7, 2)$

i	$\bar{X}_1^2$	$\bar{X}_2^2$	$\bar{X}_3^2$	CLUSTER
A	~~1,11~~	6,32	6,7	1
B	4,71	5,09	~~1~~	3
C	6,94	6,08	~~1,41~~	3
D	~~1,11~~	4,47	5	1
E	~~6,5~~	5,83	5,38	1
F	~~1,5~~	5,09	6,4	1
G	5,31	6,08	~~1,41~~	3
H	5,31	~~1,41~~	6,08	2
I	5,22	4	~~1~~	3
J	5,22	~~1~~	4	2
K	6,18	4,12	~~1,41~~	3
L	6,5	~~1~~	5,09	2

THE CLUSTERS DON'T CHANGE, THERFORE WE CAN STOP.

Scanned by CamScanner