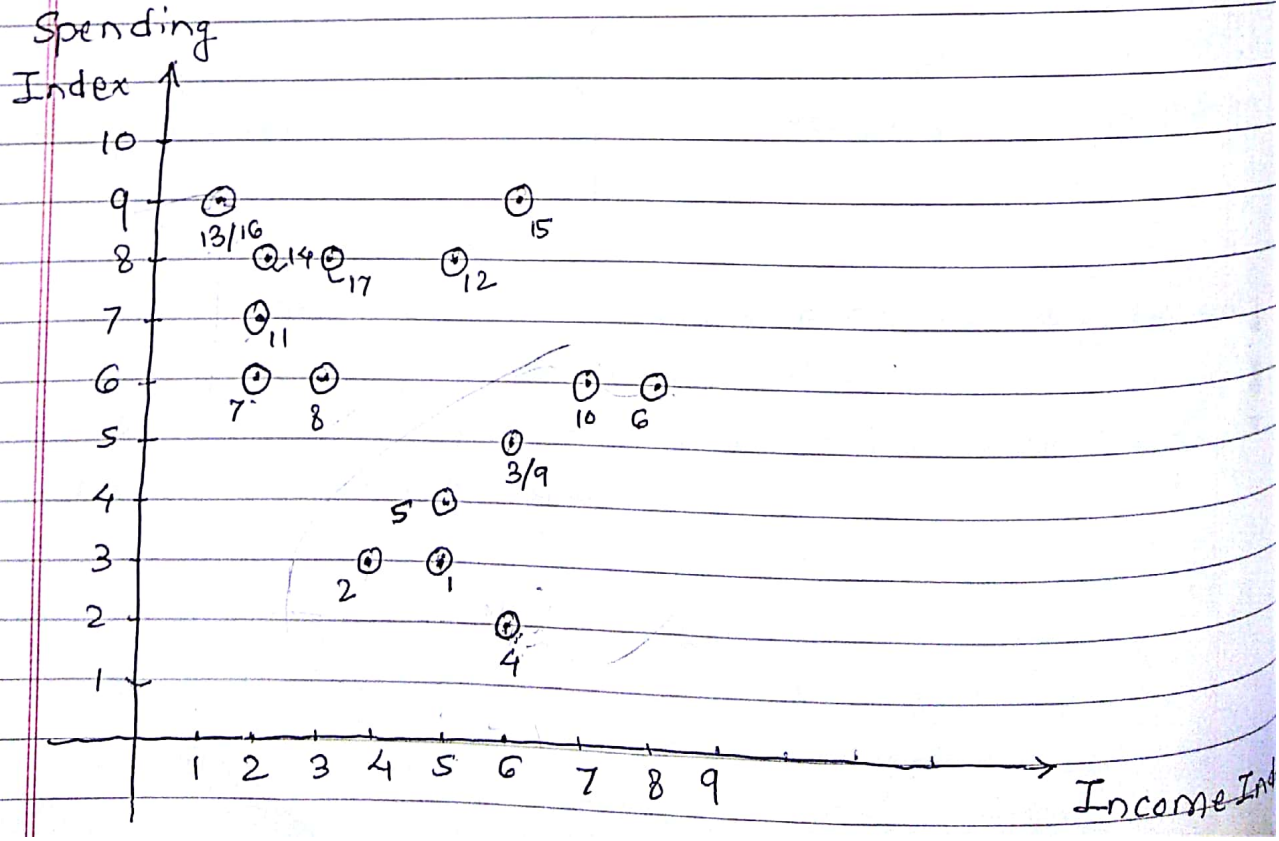
Q Develop K means Clustering

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

Shopper	Spending Index	Income Index
1	3	5
2	3	4
3	5	6
4	2	6
5	4	5
6	6	8
7	6	2
8	6	3
9	5	6
10	6	7
11	7	2
12	8	5
13	9	1
14	8	2
15	9	6
16	9	1
17	8	3



K-Means Clustering Problem

Step 1 - Initially any 2 points are considered as centroids for 1st iteration. We will consider max distance separated points as initial centroids.
By visually looking at plot we consider point 13 and point 4 as initial centroids.

Step 2 - Assigning clusters to each point.

Data Point	Cluster 1 Individual	centroid	Cluster 2 Individual	Centroid
	13	(1,9)	4	(6,2)
1(5,3)	13	(1,9)	1,4	(5.5,2.5)
2(4,8)	13	(1,9)	1,2,4	(5,2.7)
3.(6,5)	13	(1,9)	1,2,3,4	(5.5,3.25)
4(6,2)	13	(1,9)	1,2,3,4	(5.5,3.25)
5(5,4)	13	(1,9)	1,2,3,4,5	(5.2,3.4)
6(8,6)	13	(1,9)	1,2,3,4,5,6	(5.7,3.8)
7(2,6)	7,13	(1.5,7.5)	1,2,3,4,5,6	(5.7,3.8)
8(3,6)	7,8,13	(2,7)	1,2,3,4,5,6	(5.7,3.8)
9(6,5)	7,8,13	(2,7)	1,2,3,4,5,6,9	(5.7,4)
10(7,6)	7,8,13	(2,7)	1,2,3,4,5,6,9,10	(5.9,4.3)
11(2,7)	7,8,11,13	(2,7)	1,2,3,4,5,6,9,10	(5.9,4.3)
12(5,8)	7,8,11,12,13	(2.6,7.2)	1,2,3,4,5,6,9,10	(5.9,4.3)
13(1,9)	7,8,11,12,13	(2.6,7.2)	1,2,3,4,5,6,9,10	(5.9,4.3)
14(2,8)	7,8,11,12,13,14	(2.5,7.3)	1,2,3,4,5,6,9,10	(5.9,4.3)
15(6,9)	7,8,11,12,13,14,15	(3,7.6)	1,2,3,4,5,6,9,10	(5.9,4.3)
16(1,9)	7,8,11,12,13,14,15,16	(2.8,7.8)	1,2,3,4,5,6,9,10	(5.9,4.3)
17(3,8)	7,8,11,12,13,14,15,16,17	(2.8,7.78)	1,2,3,4,5,6,9,10	(5.9,4.3)

∴ Initial Centroids are

(1) (2.8, 7.78) (2) (5.9, 4.3)

Step 3 - Calculate distance of each data point from centroid of cluster

Clusters Assigned	Individual	Distance from C1 (2.8, 7.78)	Distance From C2 (5.9, 4.3)
C2	1 (5, 3)	5.26	1.58
C2	2 (4, 3)	4.92	2.30
C2	3 (6, 5)	4.23	0.70
C2	4 (6, 2)	6.60	2.30
C2	5 (5, 4)	4.37	0.94
C2	6 (8, 6)	5.50	2.70
C1	7 (2, 6)	1.95	4.25
C1	8 (3, 6)	1.79	3.36
C2	9 (6, 5)	4.23	0.70
C2	10 (7, 6)	4.56	2.02
C1	11 (2, 7)	1.11	4.74
C1	12 (5, 8)	2.21	3.80
C1	13 (1, 9)	2.17	6.78
C1	14 (2, 8)	0.82	5.37
C1	15 (6, 9)	3.42	4.70
C1	16 (1, 9)	2.17	6.78
C1	17 (3, 8)	0.29	4.70