

Assignment 9

Swapnil Asawa, swa12

1 (a)

(0,0) (0,5) (6,7) (7,0)

Initial means

(0,0) (7,0)

From (0,0): (0,0) is 0 dist away
(0,5) is 5 units away
(6,7) is 9.219 units
(7,0) is 7 units

From (7,0): (0,0) is 7 units away
(0,5) is 8.6 units
(6,7) is $\sqrt{1^2 + 7^2} = 7.07$ units away
(7,0) is 0 units away

→ New cluster near 0,0

= (0,0) (0,5)

→ Mean₁ = (0, 2.5)

cluster of (7,0)

= (6,7), (7,0)

→ Mean₂ = (6.5, 3.5)

2nd iteration → Means

Dist of pts from cluster means.

From (0, 2.5): (0,0) is 2.5 units away
(0,5) " " "
(6,7) " 6.95 units away
(7,0) " 2.43 " "

From (6.5, 3.5): (0,0) is 7.38 "
(0,5) is 6.67 "
(6,7) is 3.58 "
(7,0) is 3.56 " "

New clusters

$(0,0), (0,5), (6,7), (7,0)$

Same clusters as in prev iteration
→ Stop computing as the K-Means algo has converged

(b) Mean $3,3$ $(7,0)$

Dist from $3,3$: $(0,0)$ is 4.24
 $(0,5)$ " 3.6
 $(6,7)$ is 5
 $(7,0)$ " 5 units away.

Dist from $(7,0)$:

7
 8.6
 7.07

→ cluster 1
 $(0,0), (0,5), (6,7)$

mean = $\frac{6}{3}, \frac{12}{3}$
 $= (2, 4)$

cluster 2
 $(7,0)$

↓ mean = $(7,0)$

Iter 2: Dist

from $(2,4)$

$(0,0)$ is 4.472
 $(0,5)$ is 2.236
 $(6,7)$ is 5 units
 $(7,0)$ is 6.20 units

From $(7,0)$:

$(0,0)$ is 7
 $(0,5)$ is 8.6
 $(6,7)$ is 7.07
 $(7,0)$ is 0

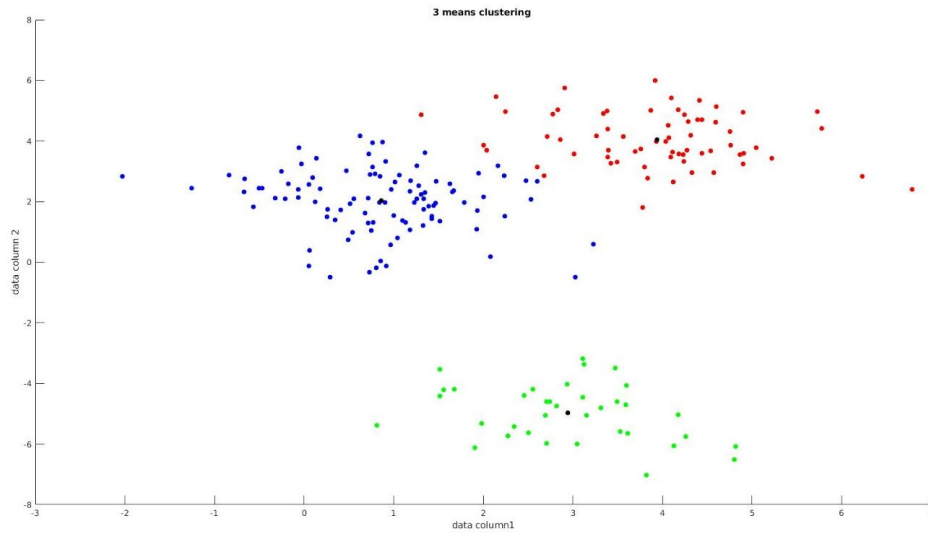
Again, same cluster 5%
cluster 1: (0,0) (0,5) (6,7) cluster 2: (7,0)
The algorithm has converged with
 $\mu_1 = (2,4)$ $\mu_2 = (7,0)$

~~mean~~

2.

a) Cluster sizes:

98, 36, 66

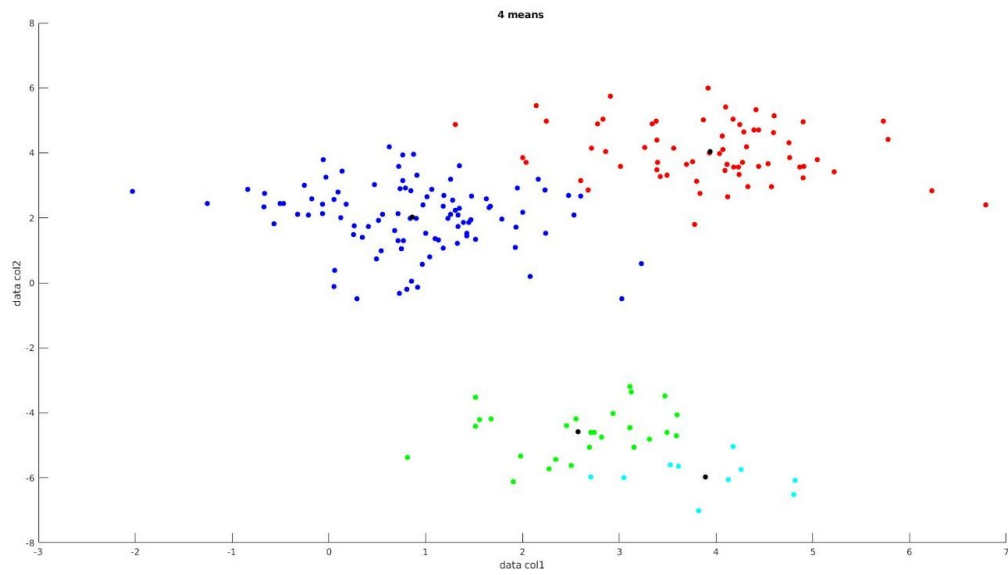


3 Means:

[0.862702047755102,2.027122220132654;
2.940840510833333,-4.969638325000001;
3.941395251515153,4.038120454545455]

b) 4 clusters

Sizes: 66,26,98,10

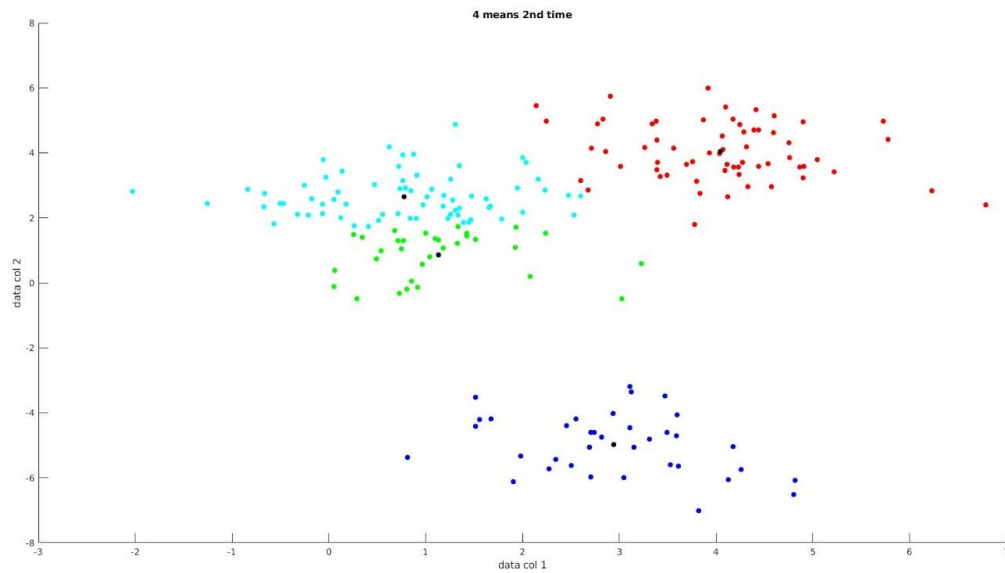


4 Means:

3.94139525151515	4.03812045454546
2.57606378807692	-4.58611246538462
0.862702047755102	2.02712222013265
3.88925999000000	-5.96680556000000

c)

sizes: 63, 32,36,69



4.04414937301588 4.03302113174603
 1.12967644646875 0.863118446031250
 2.94084051083333 -4.96963832500000
 0.778925079608696 2.65904066666667

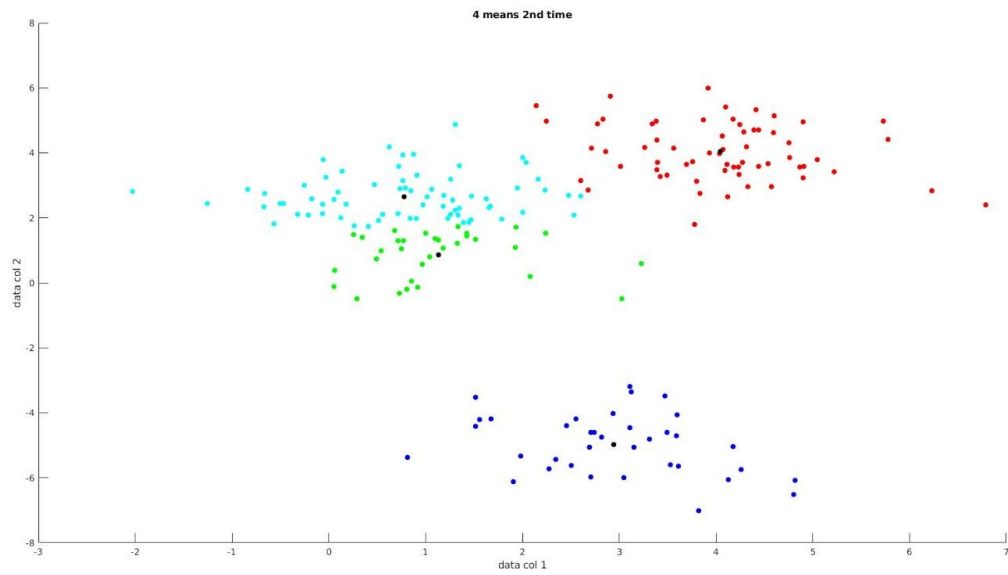
d)

I have implemented the below formula in the attached **best_clustering.m** file function

$$\min_{\mathbf{S}} \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - u_i\|^2 \quad u_i = \text{center of cluster } S_i$$

It calculates the square center point difference.

2nd cluster is the better 1st according to the function.



e)

Cluster Sizes:

66	13	23	98
96	67	22	15
63	36	52	49
66	26	98	10
96	36	40	28
85	36	53	26
98	66	23	13
98	26	10	66
63	69	36	32
69	32	63	36
29	89	46	36
10	98	26	66
98	66	26	10
98	23	66	13
39	36	97	28
36	63	40	61
98	10	26	66
98	66	26	10
66	23	13	98
69	63	36	32
65	39	36	60
40	28	36	96
26	36	87	51

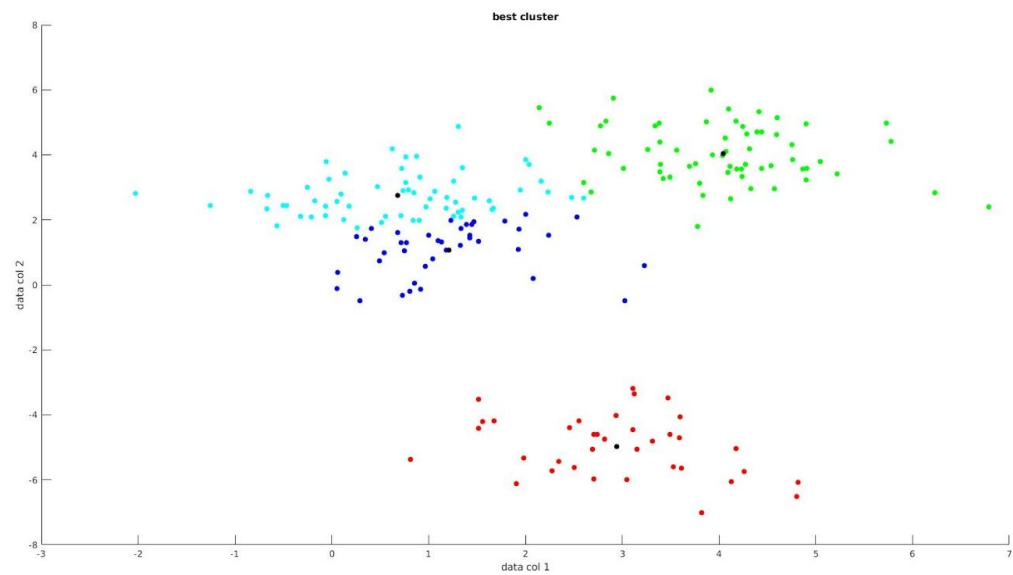
65	60	36	39
36	60	40	64
10	66	98	26
38	63	36	63
98	66	26	10
69	63	32	36
49	36	63	52

sq center point distances:

329.9973
333.8049
282.4565
329.1850
315.4905
301.5817
329.9973
329.1850
283.0650
283.0650
305.2981
329.1850
329.1850
329.9973
315.9734
281.9109
329.1850
329.1850
329.9973
283.0650
284.3597
315.4905
314.4052
284.3597
281.9975
329.1850
281.9388
329.1850
283.0650
282.4565

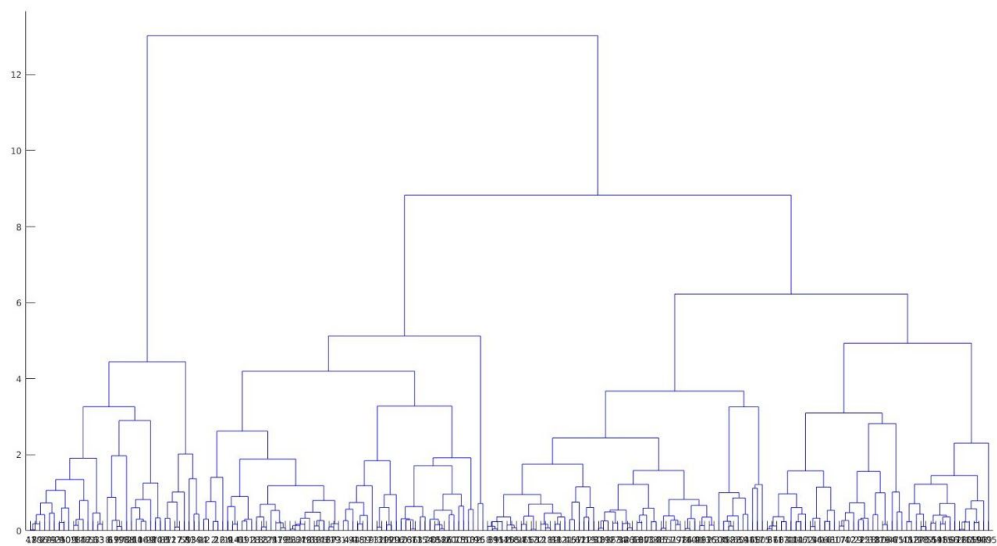
Best cluster =**16th** using the best_distance.m

It's plot:

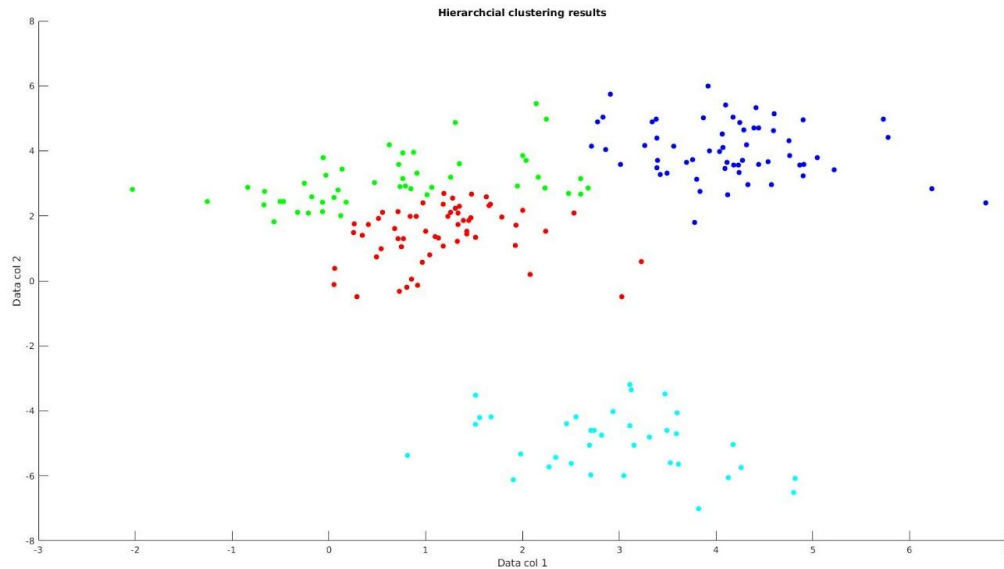


3.

a)



b)



Sizes

58,47,59,36

Centroids are calculated using means of every class which gave sq center-point **dist= 2.8229e+03** which is more than k means best results, thus k means best result is better than hierarchical clustering if we take square center point distance as our metric.

The clusters as well as cluster sizes are different.

In face none of the 30 initialised k means converged to hierarchical plot.

However hierarchical plot is near to best k means plot.