Assignment - 9 (Clustering)

Question 1: k means clustering, 2 centroids

Points: 1, 4, 9, 16, 25, 36, 49, 64, 81, 100

(36, 64) =>

1, 4, 9, 16, 25, 36, 49 | 64, 81, 100

x_1 x_2

Mean of x_1 = 20

Mean of x_2 = 81.67

Mean of (x_1, x_2) = 32.9 (to divide here for recentering)

1 4 9 16 25 | 36 49 64 81 100

Here 2 centroids are recentered from 2, 10 x_2

(9, 49)

1, 4, 9, 16, 25 | 36, 49 64 81 100

x_1 x_2

Mean of x_1 = 11

Mean of x_2 = 66

Mean of x_1 & x_2 = 38.5

Recenter of this Point (Range)

1 4 9 16 25 36 | 49 64 81 100

In this case only one point is reclassified

Question 2 L_1, L_2 norm

Centroid $c_1 = (0,0)$

Centroid $c_2 = (100,40)$

$(0,0) \Rightarrow L_1$ norm

$(100,40) \Rightarrow L_2$ norm

We find the L_2 & L_1 norm for the given option with 2 centroids. Point should be clustered.

$(52,13)$

$c_1 = (0,0)$

L_1 norm $= 65$

L_2 norm $= 53.6$

$c_2 = (100,40)$

L_1 norm $= (100-52) + (40-13)$

$= 75$

L_2 norm $= \sqrt{(48)^2 + (27)^2} = 55.07$

Smallest of L_1 is 65 which is for point $(0,0)$

Smallest of L_2 is 53.6 which is $(0,0)$

$(53,15)$

$c_1 = (0,0)$

$L_1 = |0-53| + |0-15|$

$= 68$

$L_2 = \sqrt{(53)^2 + (15)^2}$

$= 55.08$

$c_2 = (100,40)$

L_1 norm $= 72$

L_2 norm $= \sqrt{2434} = 49.33$

Smallest of L_1 is 68 which is for point $(0,0)$

Smallest of L_2 is 49.33 which is for point $(100,40)$

~~B~~ $(53,15)$ //

Question 3 Hierarchical clustering

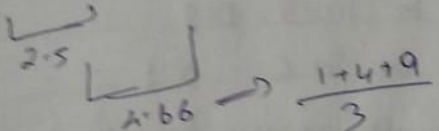
Points → (1, 4, 9, 16, 25, 36, 49, 64)

Initially there are 8 clusters. We have to move till there are 2 clusters.

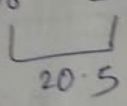
1 4 9 16 25 36 49 64

(Smallest distance between 2 points)

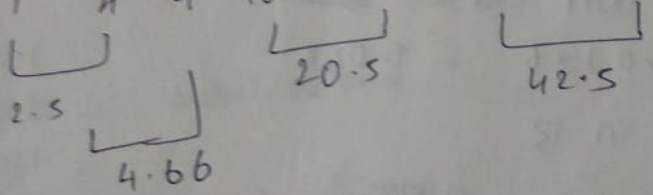
1 4 9 16 25 36 49 64



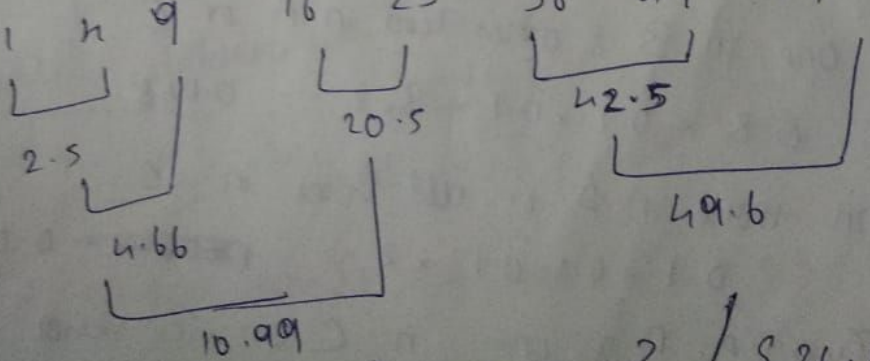
1 4 9 16 25 36 49 64



1 4 9 16 25 36 49 64

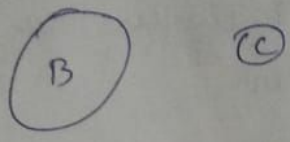


1 4 9 16 25 36 49 64



The clusters are {1, 4, 9, 16, 25} / {36, 49, 64}

Question 4:



radius = 2

B → (0,0) → 8,000 points

A → (-10,0) → 1000 points

C → (10,0) → 1000 points

Probability of point to be in A = 0.1

probability of point to be in B = 0.8

probability of point to be in C = 0.1

There are 27 ways for points x, y, z to be in A, B, C.

① Two in B & one in C

$0.8 \times 0.8 \times 0.1 \times 3 \times 2 = 0.384$

② All three in diff clusters

$0.8 \times 0.1 \times 0.1 \times 6 = 0.048$

③ All three in B

$0.8 \times 0.8 \times 0.8 = 0.512$

④ One in B & other two in A or C

$0.8 \times 0.1 \times 0.1 \times 3 \times 2 = 0.048$

⑤ All three in A or all three in C

$0.1 \times 0.1 \times 0.1 \times 2 = 0.002$

⑥ Two in A & one in C (or) vice-versa

$0.1 \times 0.1 \times 0.1 \times 2 \times 3 = 0.006$

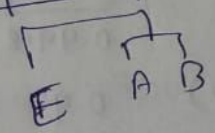
Question 5: Hierarchical clustering

A = (0,0), B = (10,10), E = (21,21)

D = (33,33), E = (5,7), F = (28,6)

Calculate distance between each pairs

	A	B	C	D	E	F
A	0	14.0	29.6	46.6	27.4	28.6
B		0	15.5	32.5	17.7	18.4
C			0	16.9	17	16.5
D				0	88.6	27.4
E					0	31.1
F						0



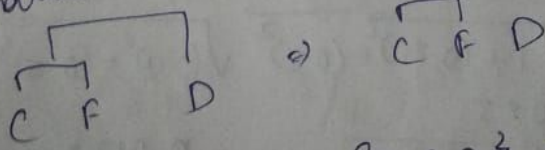
ABC

ABE => lowest distance

ABF =

ABD => largest distance

Smallest difference with D is C



The 2 clusters are {E, A, B} and {C, F, D}