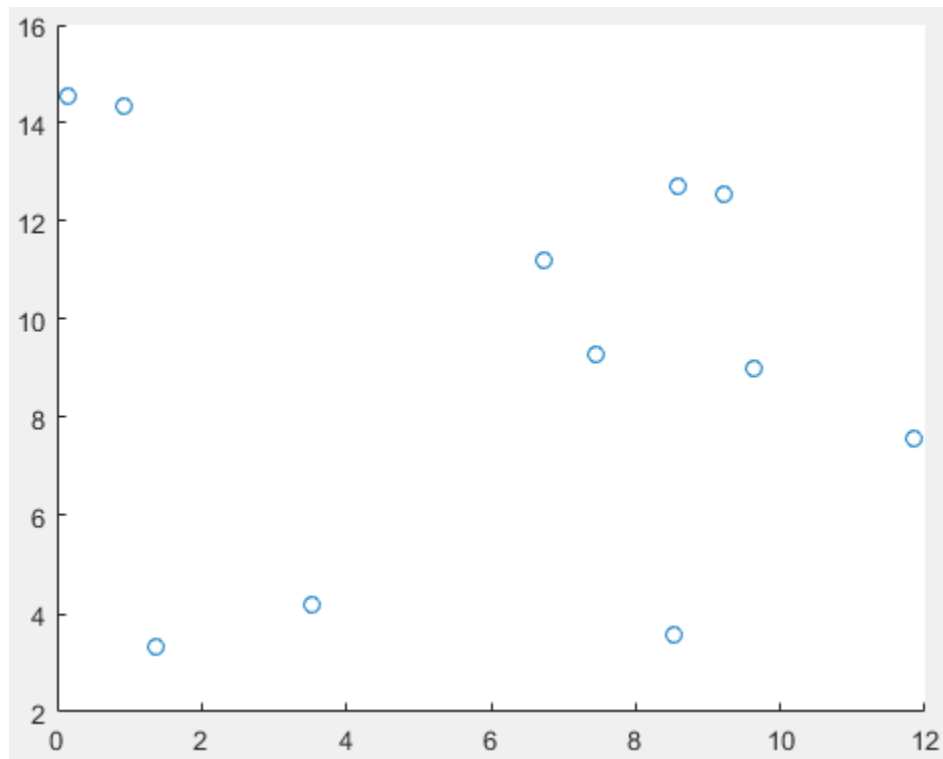


ALI HELMUT

1301154246

LOAD THE SELECTED DATA SET. VISUALIZE ALL DATA POINTS USING SCATTER PLOT IN ONE COLOR (NO NEED TO GIVE DIFFERENT COLOR FOR EACH CLASS). USE ATTRIBUTE 1 AS X - AXIS, ATTRIBUTE 2 AS Y -AXIS.[4 POINTS]



CREATE 4 FUNCTIONS TO MEASURE DISSIMILARITY BETWEEN 2 CLUSTERS. EACH FUNCTION APPLYING EACH OF FOLLOWING PROXIMITY MEASURES,

I. SINGLE-LINK (MIN) [6 POINTS]

Saat jenis = 1

```
function [ output_args ] = MinMaxAvgCtr( cluster1 , cluster2 , jenis)
[panjang1 lebar1] = size(cluster1);
[panjang2 lebar2] = size(cluster2);
data = [];
if(jenis < 4)
    for i=1:panjang1
        for j=1:panjang2
            data = [data; i j norm(cluster1(i,:)-cluster2(j,:))];
        end
    end
    if(jenis == 1)
        data = sortrows(data,3);
        output_args = data(1,3);
    elseif (jenis == 2)
        data = sortrows(data,3);
        output_args = data(end,3);
    elseif (jenis == 3)
        output_args = mean(data(:,3));
    end
else
    avgCluster1 = mean(cluster1);
    avgCluster2 = mean(cluster2);
    output_args = norm(avgCluster1-avgCluster2);
end
```

II. COMPLETE-LINK (MAX) [6 POINTS]

Saat jenis = 2

```
function [ output_args ] = MinMaxAvgCtr( cluster1 , cluster2 , jenis)
[panjang1 lebar1] = size(cluster1);
[panjang2 lebar2] = size(cluster2);
data = [];
if(jenis < 4)
    for i=1:panjang1
        for j=1:panjang2
            data = [data; i j norm(cluster1(i,:)-cluster2(j,:))];
        end
    end
    if(jenis == 1)
        data = sortrows(data,3);
        output_args = data(1,3);
    elseif (jenis == 2)
        data = sortrows(data,3);
        output_args = data(end,3);
    elseif (jenis == 3)
        output_args = mean(data(:,3));
    end
else
    avgCluster1 = mean(cluster1);
    avgCluster2 = mean(cluster2);
    output_args = norm(avgCluster1-avgCluster2);
end
```

III. GROUP AVERAGE [6 POINTS]

Saat jenis = 3

```
function [ output_args ] = MinMaxAvgCtr( cluster1 , cluster2 , jenis)
[panjang1 lebar1] = size(cluster1);
[panjang2 lebar2] = size(cluster2);
data = [];
if(jenis < 4)
    for i=1:panjang1
        for j=1:panjang2
            data = [data; i j norm(cluster1(i,:)-cluster2(j,:))];
        end
    end
    if(jenis == 1)
        data = sortrows(data,3);
        output_args = data(1,3);
    elseif (jenis == 2)
        data = sortrows(data,3);
        output_args = data(end,3);
    elseif (jenis == 3)
        output_args = mean(data(:,3));
    end
else
    avgCluster1 = mean(cluster1);
    avgCluster2 = mean(cluster2);
    output_args = norm(avgCluster1-avgCluster2);
end
```

IV. CENTROID-BASED [6 POINTS]

Saat jenis = 1

```
function [ output_args ] = MinMaxAvgCtr( cluster1 , cluster2 , jenis)
[panjang1 lebar1] = size(cluster1);
[panjang2 lebar2] = size(cluster2);
data = [];
if(jenis < 4)
    for i=1:panjang1
        for j=1:panjang2
            data = [data; i j norm(cluster1(i,:)-cluster2(j,:))];
        end
    end
    if(jenis == 1)
        data = sortrows(data,3);
        output_args = data(1,3);
    elseif (jenis == 2)
        data = sortrows(data,3);
        output_args = data(end,3);
    elseif (jenis == 3)
        output_args = mean(data(:,3));
    end
else
    avgCluster1 = mean(cluster1);
    avgCluster2 = mean(cluster2);
    output_args = norm(avgCluster1-avgCluster2);
end
```

INPUTS FOR EACH FUNCTION ARE 2 CLUSTERS CONTAINING DATA POINTS ASSIGNED TO EACH CLUSTER, WHILE THE OUTPUT IS DISSIMILARITY VALUE BETWEEN THE 2 CLUSTERS.

CREATE A FUNCTION TO IMPLEMENT HIERARCHICAL CLUSTERING WITH AGGLOMERATIVE APPROACH THAT CALLS FUNCTIONS CREATED IN POINT 2(A)I UNTIL 2(A)IV. [10 POINTS]

```
dataTrain = csvread('Hierarchical_7.csv');
scatter(dataTrain(:,1),dataTrain(:,2));

jenisProximity = 4; % 1 = min , 2 = max , 3 = avg , 4 = centroid
dataTrainCluster = [];
for i=1 : length(dataTrain)
    dataTrainCluster = [dataTrainCluster; i , dataTrain(i,:) ];
end
dataTrainClusterBaru = [];
banyakClass = max(dataTrainCluster(:,1));
for z=1 : length(dataTrain)-1
    jarakAntarCluster = [];

    length(unique(dataTrainCluster(:,1)))
    for i=1 : banyakClass-1
        for k=i+1 : banyakClass
            cluster1 = [];
            cluster2 = [];
            for j=1 : length(dataTrainCluster)
                if(dataTrainCluster(j,1) == i)
                    cluster1 = [cluster1 ; dataTrain(j,:)];
                elseif (dataTrainCluster(j,1) == k)
                    cluster2 = [cluster2 ; dataTrain(j,:)];
                end
            end
            if((isempty(cluster1) | isempty (cluster2))~=1)
                jarakAntarCluster = [jarakAntarCluster; i k
MinMaxAvgCtr(cluster1,cluster2,jenisProximity)];
            end
        end
    end
    jarakAntarCluster = sortrows(jarakAntarCluster,3);
    dataTrainCluster2 = [];
    banyakClass = banyakClass +1;
    for i=1 : length(dataTrain)
        if( dataTrainCluster(i,1) == jarakAntarCluster(1,1) |
dataTrainCluster(i,1) == jarakAntarCluster(1,2))
            dataTrainClusterBaru = [dataTrainClusterBaru; banyakClass
dataTrainCluster(i,1), jarakAntarCluster(1,3)];
            dataTrainCluster2 = [dataTrainCluster2; banyakClass ,
dataTrain(i,:)];
        else
            dataTrainCluster2 = [dataTrainCluster2; dataTrainCluster(i,:)];
        end
    end
    dataTrainCluster = dataTrainCluster2;
end
```

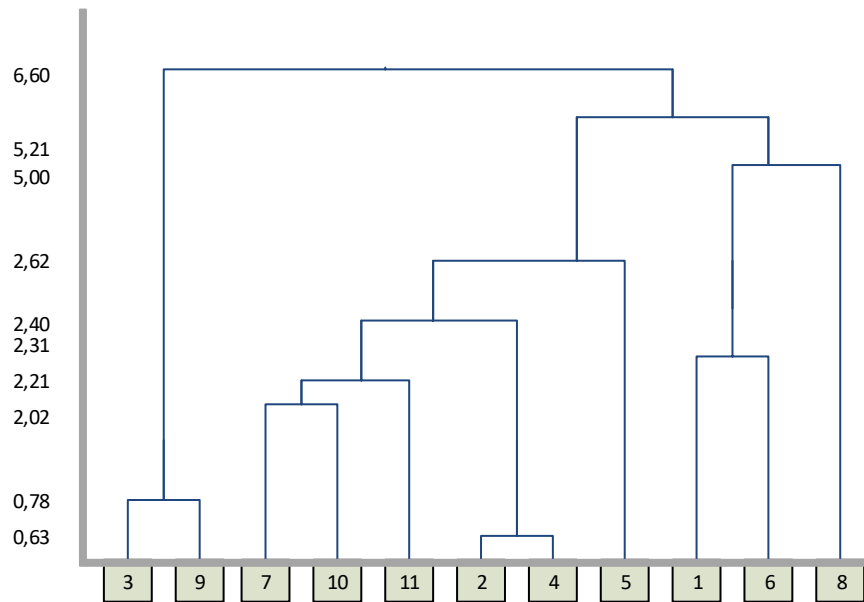
VISUALIZE EACH CLUSTERING RESULTS AS DENDROGRAM (TREE DIAGRAM) AND NESTED CLUSTERS. WE'LL HAVE 4 PAIRS OF VISUALIZATION OF DENDROGRAM AND NESTED CLUSTERS. [8 POINTS]

MATRIX HASIL OUTPUT PROGRAM :

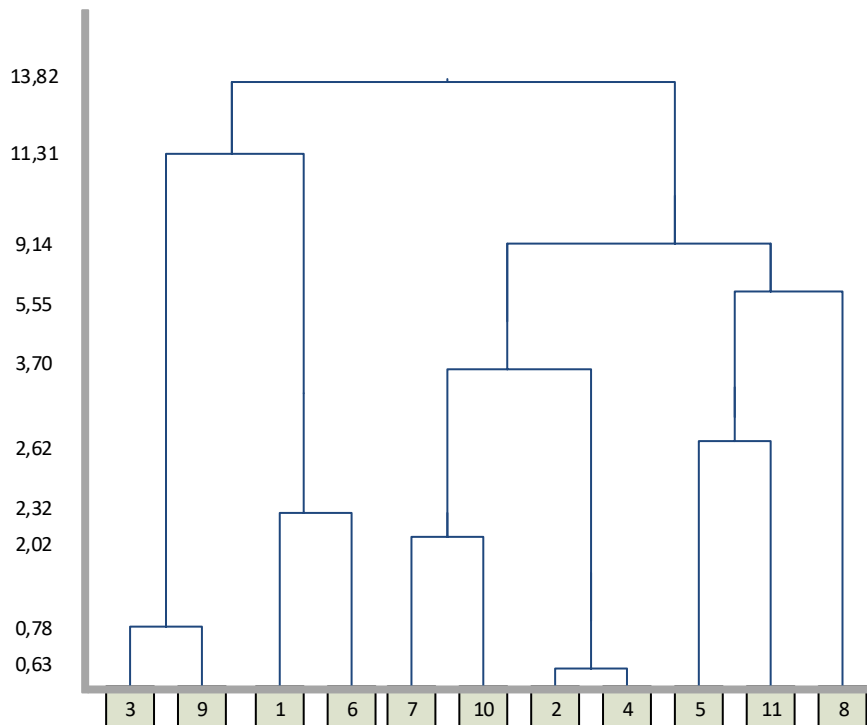
Single Link			complete link			group average			centroid base		
level	base	range	level	base	range	level	base	range	level	base	range
12	2	0.6361	12	2	0.6361	12	2	0.6361	12	2	0.24025
12	4	0.6361	12	4	0.6361	12	4	0.6361	12	4	0.24025
13	3	0.78388	13	3	0.78388	13	3	0.78388	13	3	0.27516
13	9	0.78388	13	9	0.78388	13	9	0.78388	13	9	0.27516
14	7	2.02869	14	7	2.02869	14	7	2.02869	14	7	0.35965
14	10	2.02869	14	10	2.02869	14	10	2.02869	14	11	0.35965
15	14	2.21085	15	1	2.32439	15	1	2.32439	15	5	1.3439
15	11	2.21085	15	6	2.32439	15	6	2.32439	15	10	1.3439
16	1	2.32439	16	5	2.62468	16	5	2.62468	16	1	1.51095
16	6	2.32439	16	11	2.62468	16	11	2.62468	16	6	1.51095
17	12	2.40211	17	12	3.70501	17	12	3.13353	17	15	2.21164
17	15	2.40211	17	14	3.70501	17	14	3.13353	17	14	2.21164
18	17	2.62468	18	8	5.55868	18	17	4.49145	18	12	3.3518
18	5	2.62468	18	16	5.55868	18	16	4.49145	18	17	3.3518
19	16	5.04924	19	17	9.14219	19	15	6.1104	19	16	4.26622
19	8	5.04924	19	18	9.14219	19	8	6.1104	19	8	4.26622
20	19	5.21692	20	15	11.3117	20	19	8.7195	20	19	8.03688
20	18	5.21692	20	13	11.3117	20	18	8.7195	20	18	8.03688
21	20	6.60481	21	20	13.8353	21	20	10.2212	21	20	9.34166
21	13	6.60481	21	19	13.8353	21	13	10.2212	21	13	9.34166

DENDROGRAM :

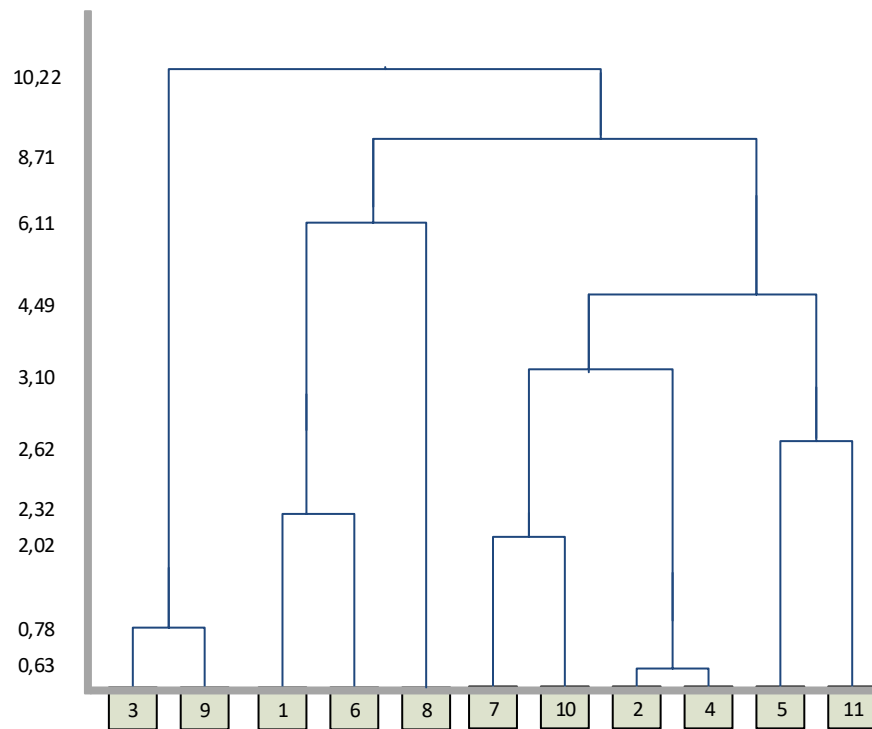
SINGLE LINK



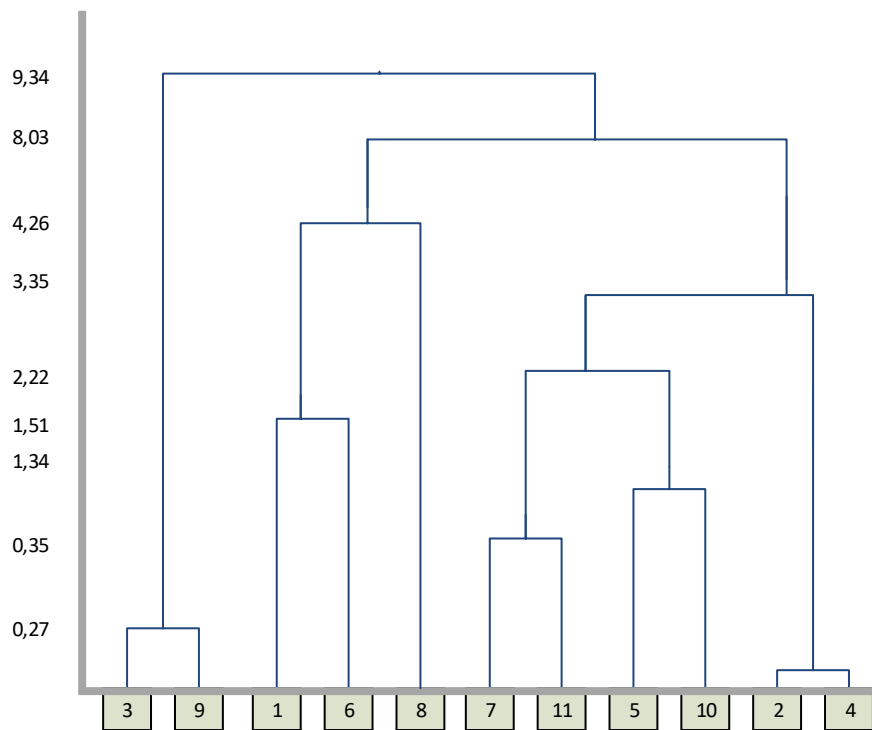
COMPLETE LINK



GROUP AVERAGE

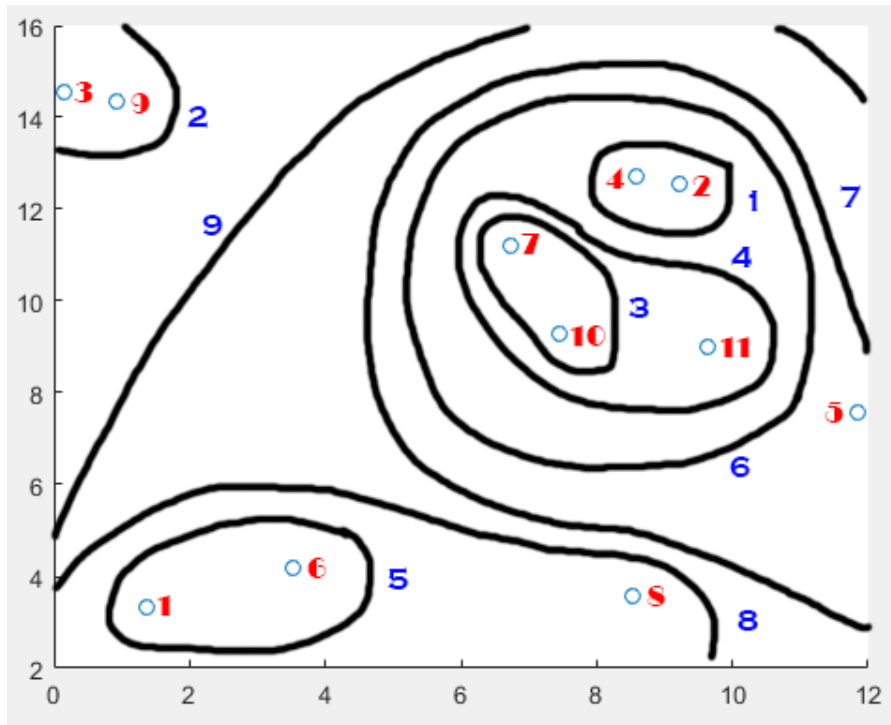


CENTROID BASE

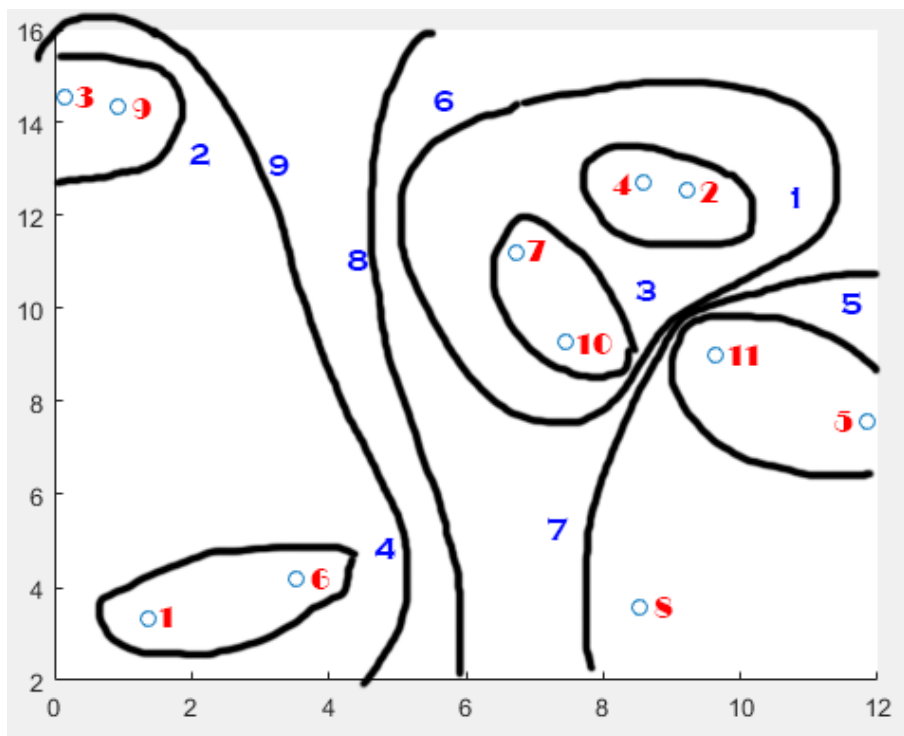


NESTED CLUSTERS

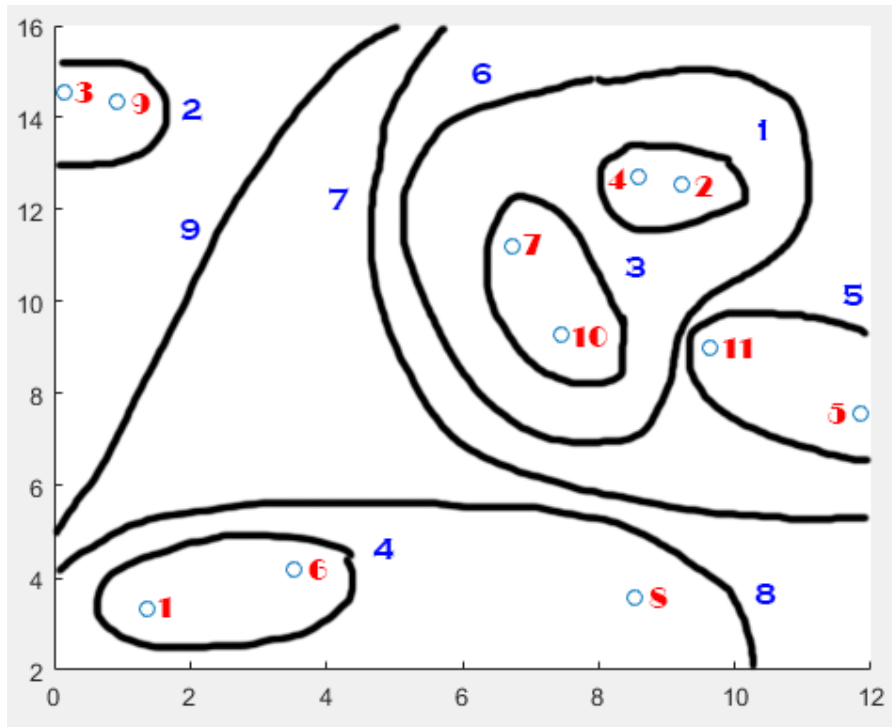
SINGLE LINK



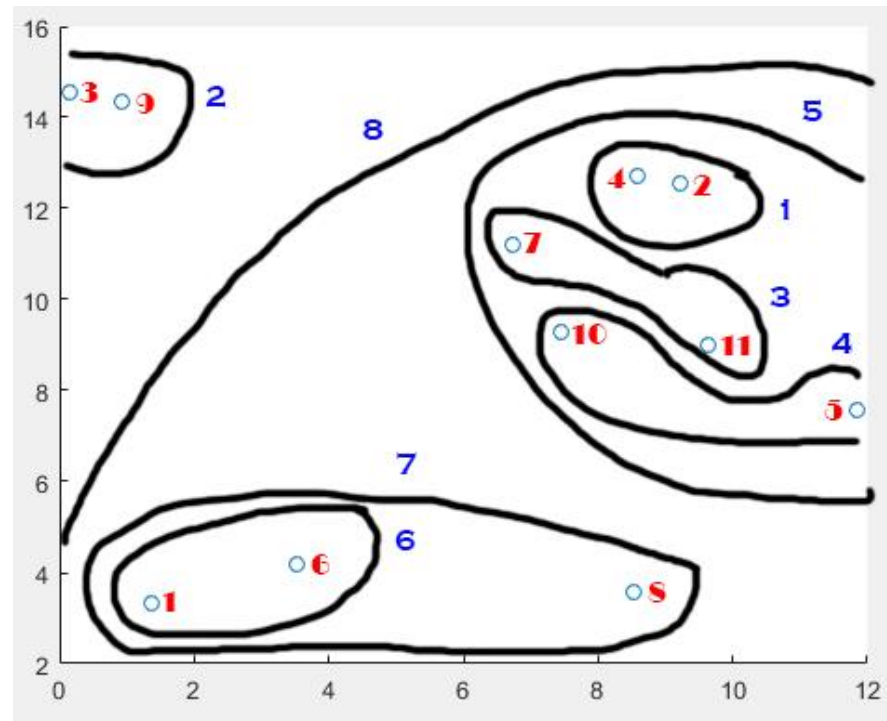
COMPLETE LINK



AREA AVERAGE



CLUSTER BASE



BY VISUALLY COMPARING CLUSTERING RESULTS CREATED FROM POINT 2(D), WHAT DO YOU THINK OF THE CLUSTERING RESULTS? GIVE EXPLANATIONS. [4 POINTS]

Setiap proximity memiliki karakteristik tersendiri, dan karakteristik tersebut berdampak pada hasil clustering. Proximity single link membuat cluster terbentuk dari jarak terkecil dari jarak minimal objek antar cluster. Proximity complete link membuat cluster terbentuk dari jarak terkecil dari jarak maksimal objek antar cluster. Proximity area average membuat cluster terbentuk dari jarak terkecil dari jarak rata rata objek antar cluster. Dan centroid base membuat cluster terbentuk dari jarak minimal centroid antar cluster.