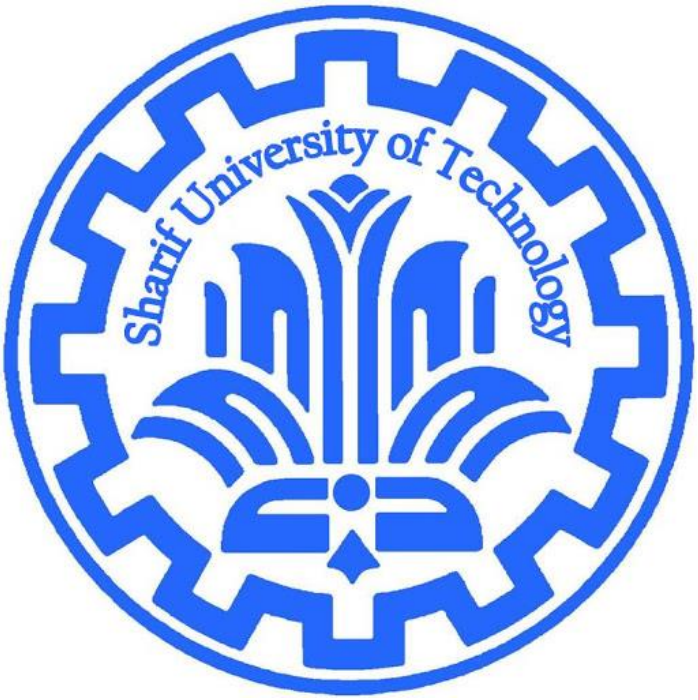
In the name of God



Computational Intelligence Computer HW2 Ex2

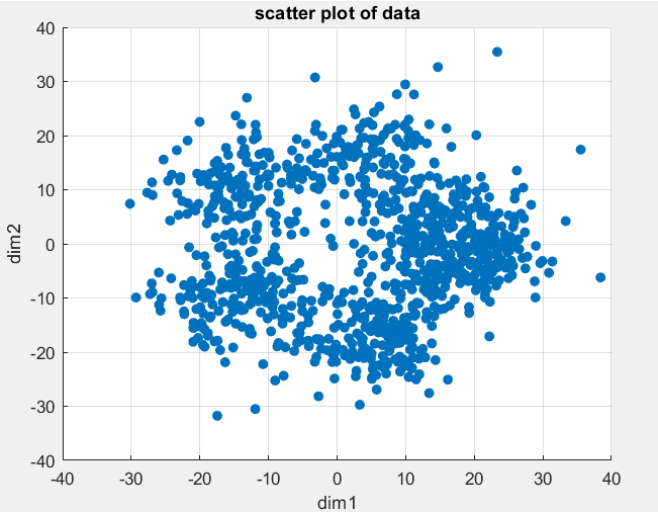
Ahmadreza Tavana: 98104852

Professor: Dr. Hajipour

Question 2

Part a

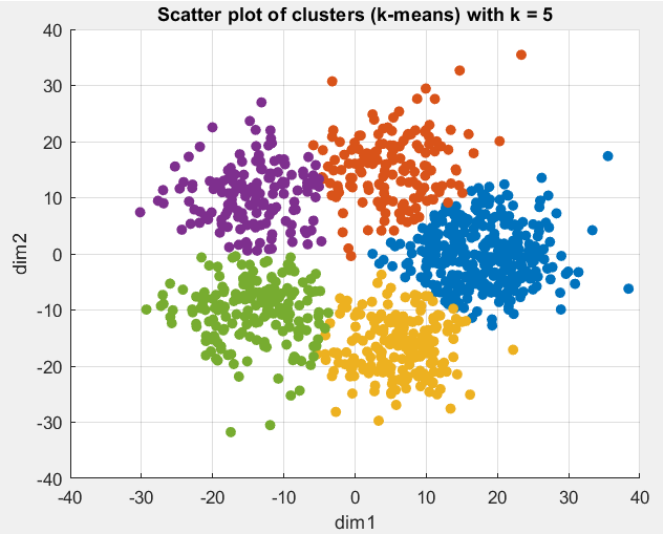
At first I show the distribution of data in a 2D scatter plot.

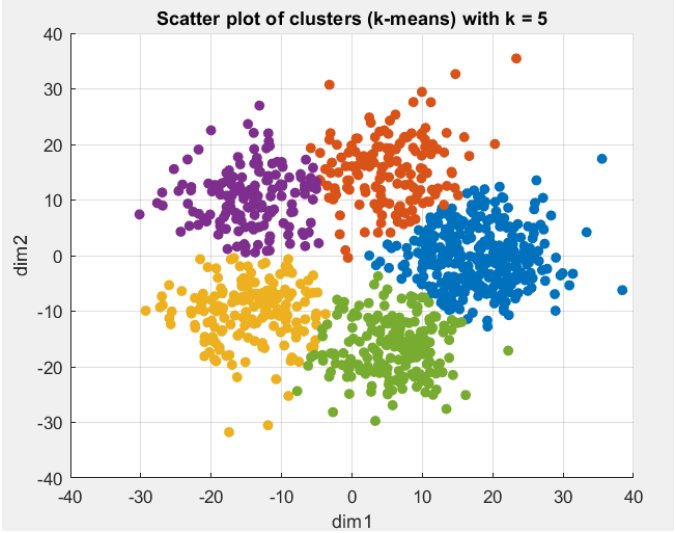


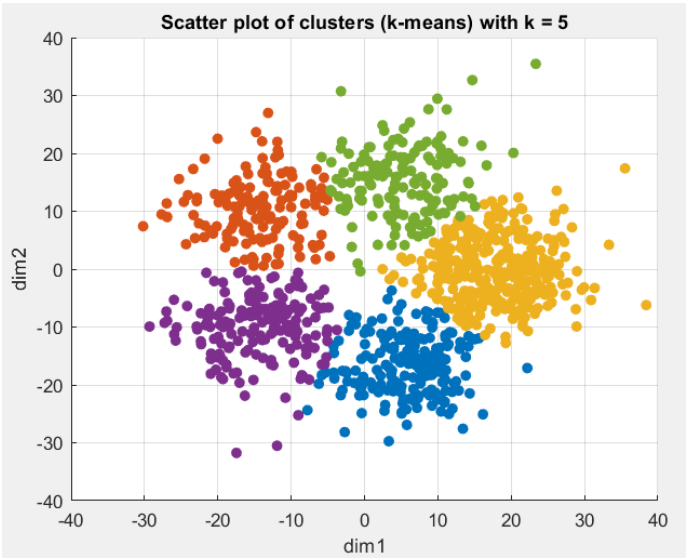
For the k-means function that I wrote, the function gets input, number of clusters, and center of cluster at starting status. Then the output is a vector of n input that n is the number of data and each of them has the value of the value of clusters.

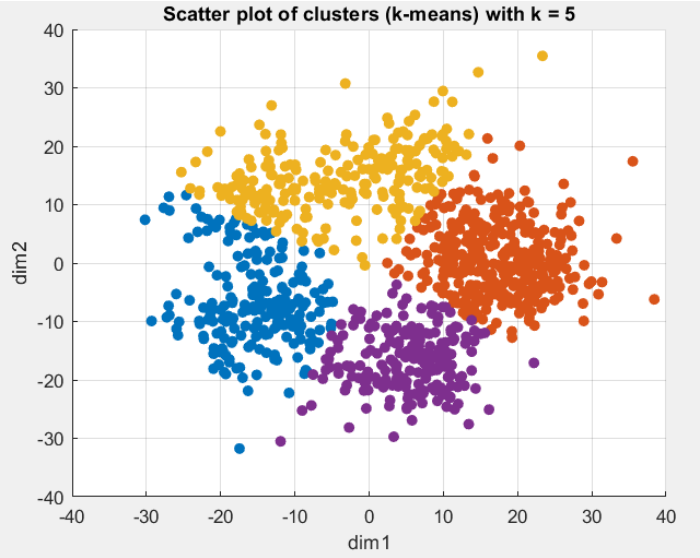
Part b

I tested my program for 4 different values of starting point and I put the results in below:

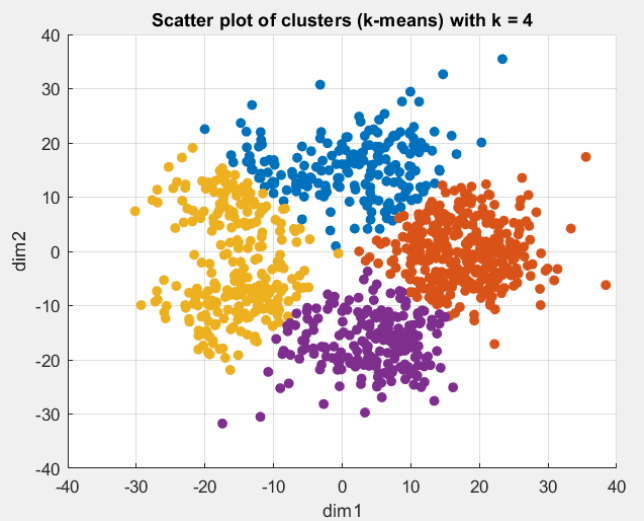


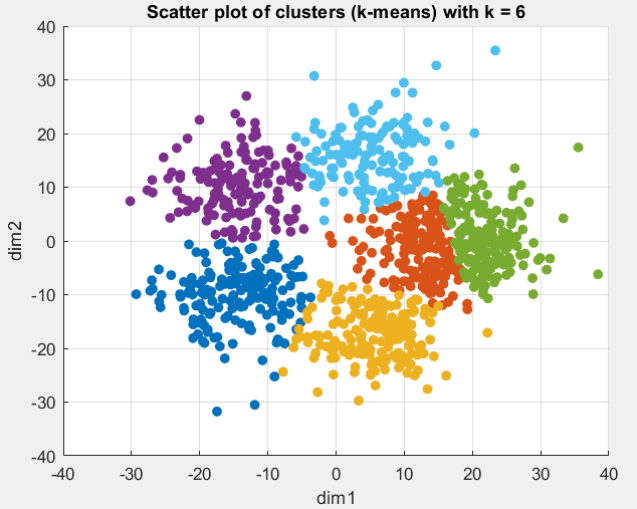

Scatter plot of clusters (k-means) with k = 5
dim2
dim1
40
30
20
10
0
-10
-20
-30
-40
-40
-30
-20
-10
0
10
20
30
40


Scatter plot of clusters (k-means) with k = 5
dim2
dim1
40
30
20
10
0
-10
-20
-30
-40
-40
-30
-20
-10
0
10
20
30
40


Scatter plot of clusters (k-means) with k = 5
dim2
dim1
40
30
20
10
0
-10
-20
-30
-40
-40
-30
-20
-10
0
10
20
30
40

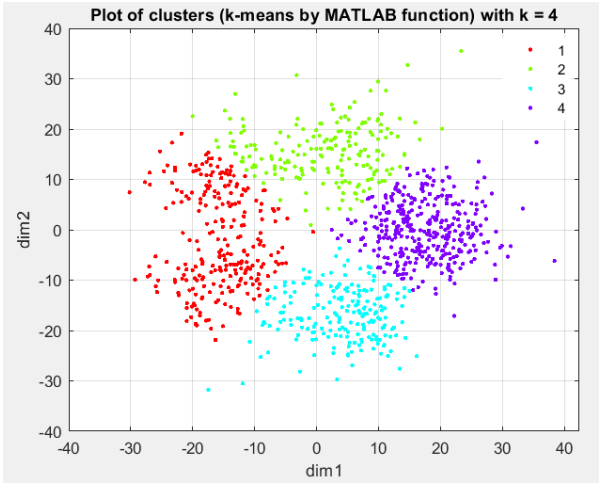
Part c

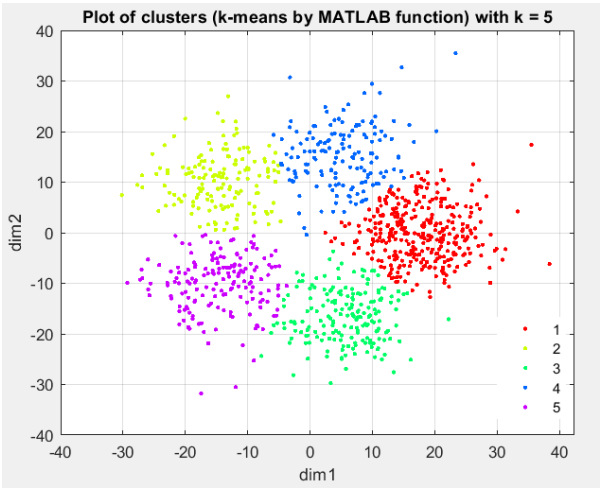


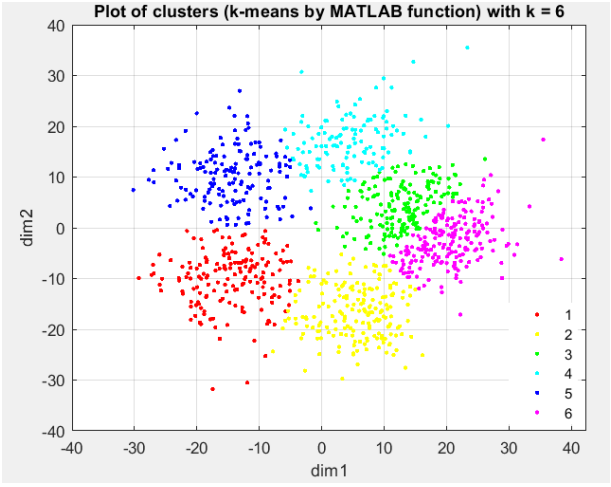


As we can see in clustering with k = 6, one of the main clusters in last part which was k = 5, is separated to two different groups and for k = 4, two of the main clusters, combined together and made one group.

Part d







As we can see, they worked as the same as my own written function for k-means clustering.

Part e

hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types:

- Agglomerative: This is a "bottom-up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- Divisive: This is a "top-down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

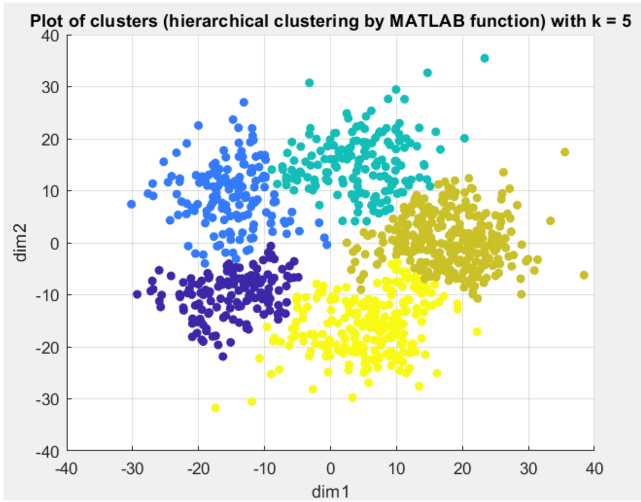
The linkage criterion determines the distance between sets of observations as a function of the pairwise distances between observations in this algorithm that I also used it in my code.

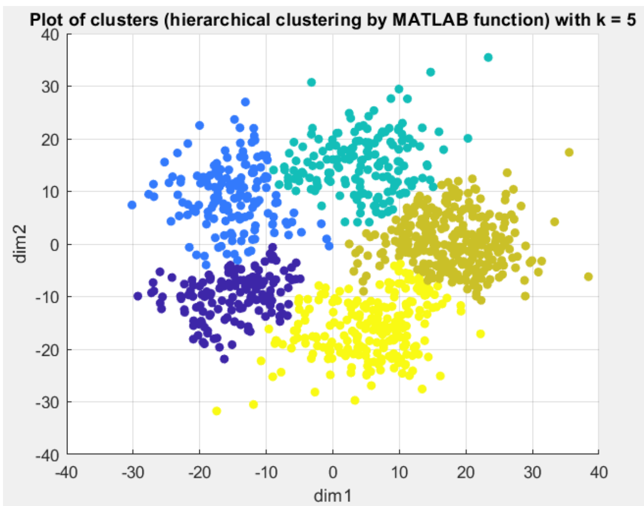
Reference: https://en.wikipedia.org/wiki/Hierarchical_clustering

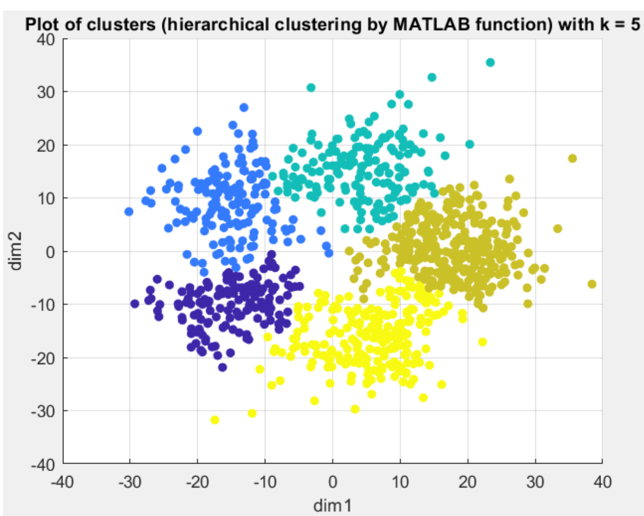
Now I put my code in below and see the result of that.

```
%% Question 2 Part e
%%%%% in this part I use hierarchical clustering with linkage and cluster
%%%%%(I explain the hierarchical clustering in my report for this part)
figure
Z = linkage(data','ward');
c = cluster(Z,'Maxclust',k);
scatter(data(1,:),data(2,:),30,c,'filled')
xlabel('dim1')
ylabel('dim2')
title(sprintf('Plot of clusters (hierarchical clustering by MATLAB function) with k = %d',k))
grid on;
```

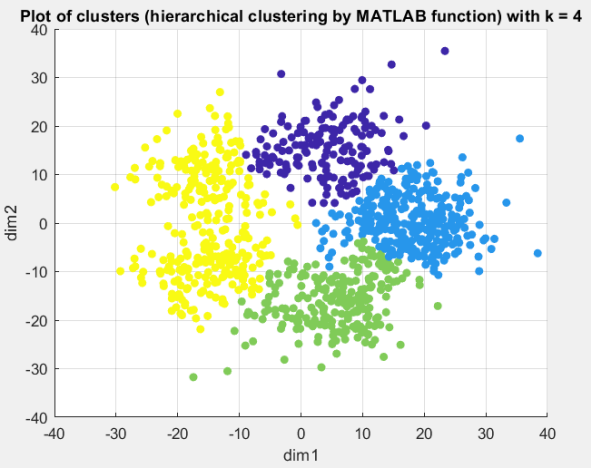
At first I try clustering with different starting status and see the results. We can see that they are all the same.

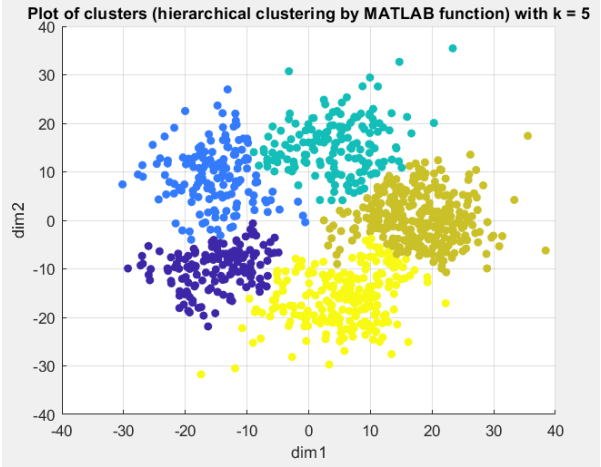


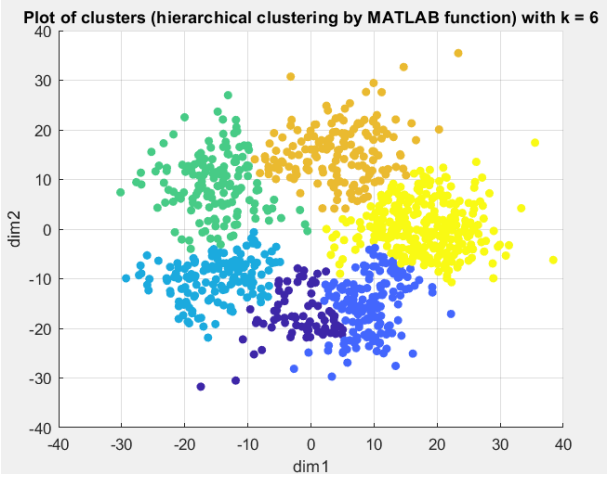




Now I try it for different values of k and see the results. We can see that the results are almost the same as k-means.







Part f

I put the results in below and as we can see the hierarchical clustering did better than k-means clustering for this data.