Sharif University of Technology
Electrical Engineering School

Computational Intelligence CHW2

RBF NETWORKS, MULTI LAYER PERCEPTRON, CLUSTRING

Armin Panjehpour
arminpp1379@gmail.com

Supervisor(s): Dr. Hajipour
Sharif University, Tehran, Iran

13/05/2022

Part.2 - Clustering using k-means & Agglomerative Clustering

Here we use k-means and Agglomerative clustering which is a type of hierarchical clustering algorithm.

Part.2.a - k-means Clustering Function

Here, I've implemented a function, k-means, which gets the data, number of clusters and the initial centroids as input and gives us the label of each sample as the output.

Here's a good overview of the k-means algorithm:

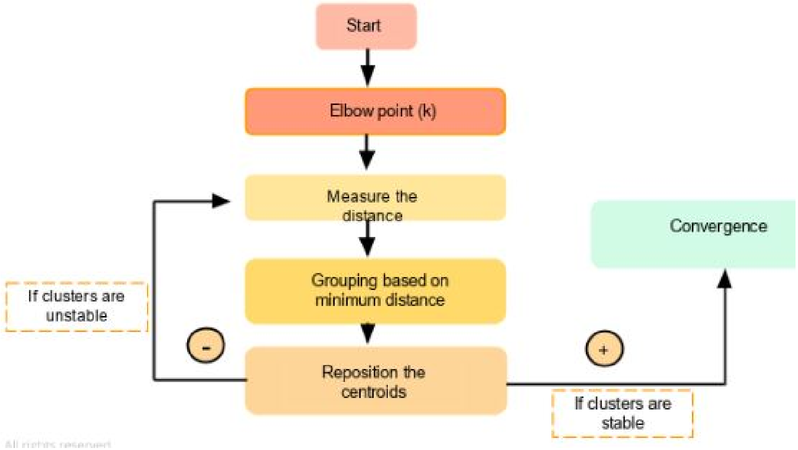

Figure 1: k-means Algorithm

```
def k_means(input_data, cluster_numbers, initial_centers):
```

Part.2.b - Apply K-means on Data with 5 Clusters

We apply the implemented function with 5 clusters on data and here's the result:

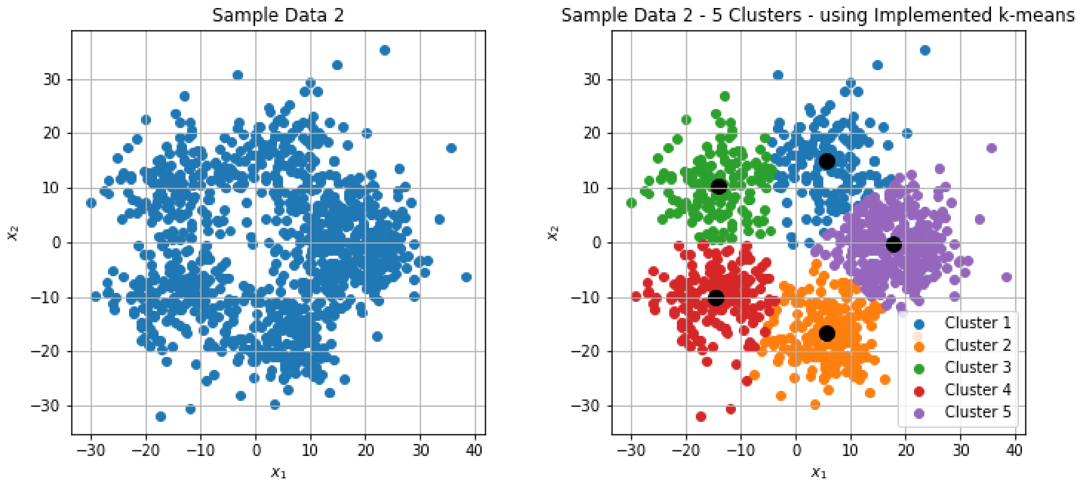

Figure 2: Sample Data 2 & Implemented k-means with 5 clusters on it

Initial centroids are created completely randomly and as many times as we run the code, the clustering will be done with high accuracy. May be there will be a little difference on clustering in each run, for example, a specific sample may change it‘s cluster to the neighbor cluster.

Part.2.c - Apply K-means on Data with 4 & 6 Clusters

We repeat the previous part but with 4 and 6 clusters, and here‘s the result:

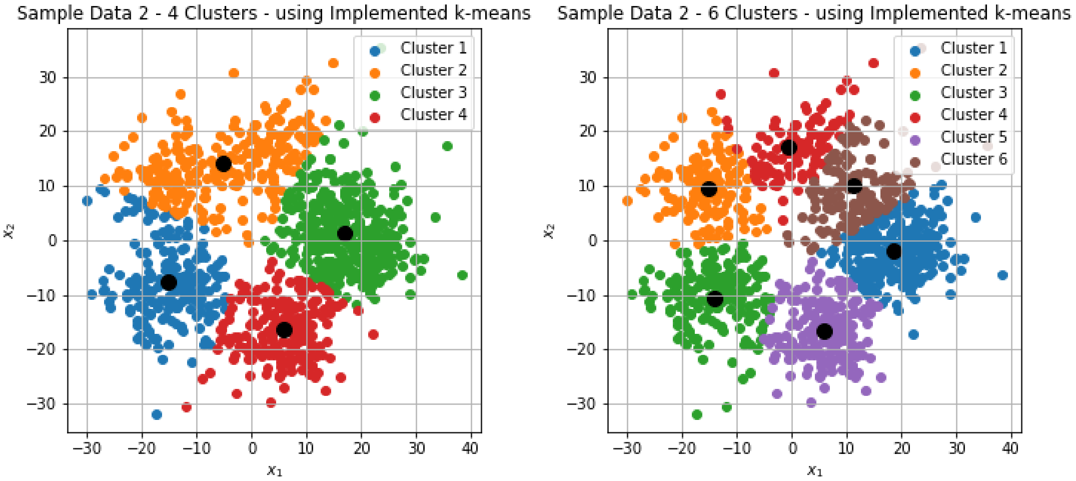

Figure 3: Implemented k-means with 4 & 6 clusters

Comparing to 5 clusters, it seems separating this data to 5 clusters is the optimum choice.

Part.2.d - Clustering Using Scikit Learn k-means

Here, we repeat the previous parts, with the internal function of the scikit learn:

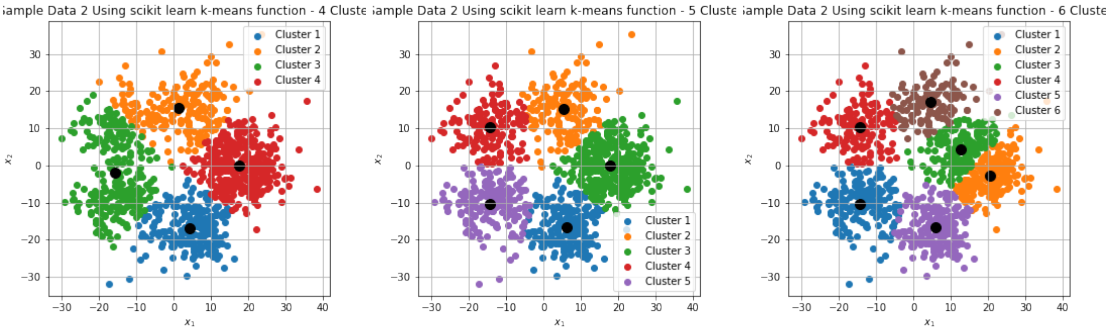

Figure 4: Scikit learn k-means with 4 & 5 & 6 clusters

As you can see, the general behavior of the this function is like the function we implemented, specially in 5 clusters which we claimed that is the optimum number of clusters.

Part.2.e - Clustering Using Agglomerative Algorithm

Agglomerative algorithm is a type of hierarchical clustering algorithm. Here we apply this algorithm on our data using sklearn library and here are the results:

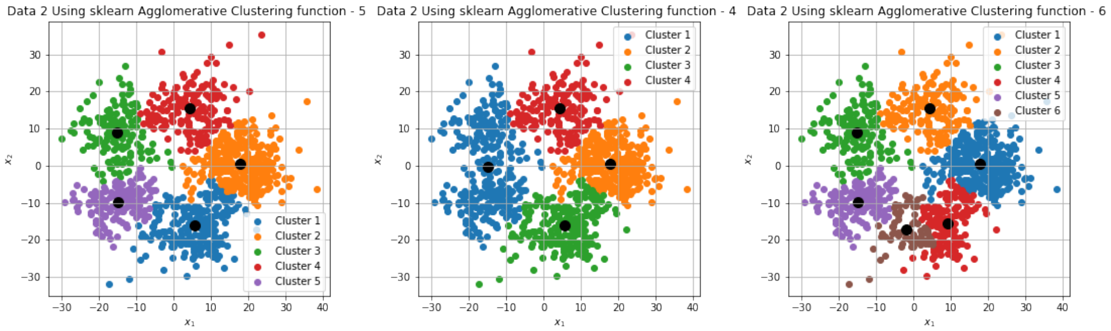

Figure 5: Agglomerative Clustering with 4 & 5 & 6 clusters

Again as you can see, the general behavior of the this function is like the previous parts, specially in 5 clusters which we claimed that is the optimum number of clusters. In 4 and 6 clusters there is a little bit of randomness in clustering that make the clusters a little different but the general behavior is the same.

Part.2.f - Repeat Previous Parts on Sample Data 3

Here we do all the previous parts on Sample Data 3 with 3 clusters:

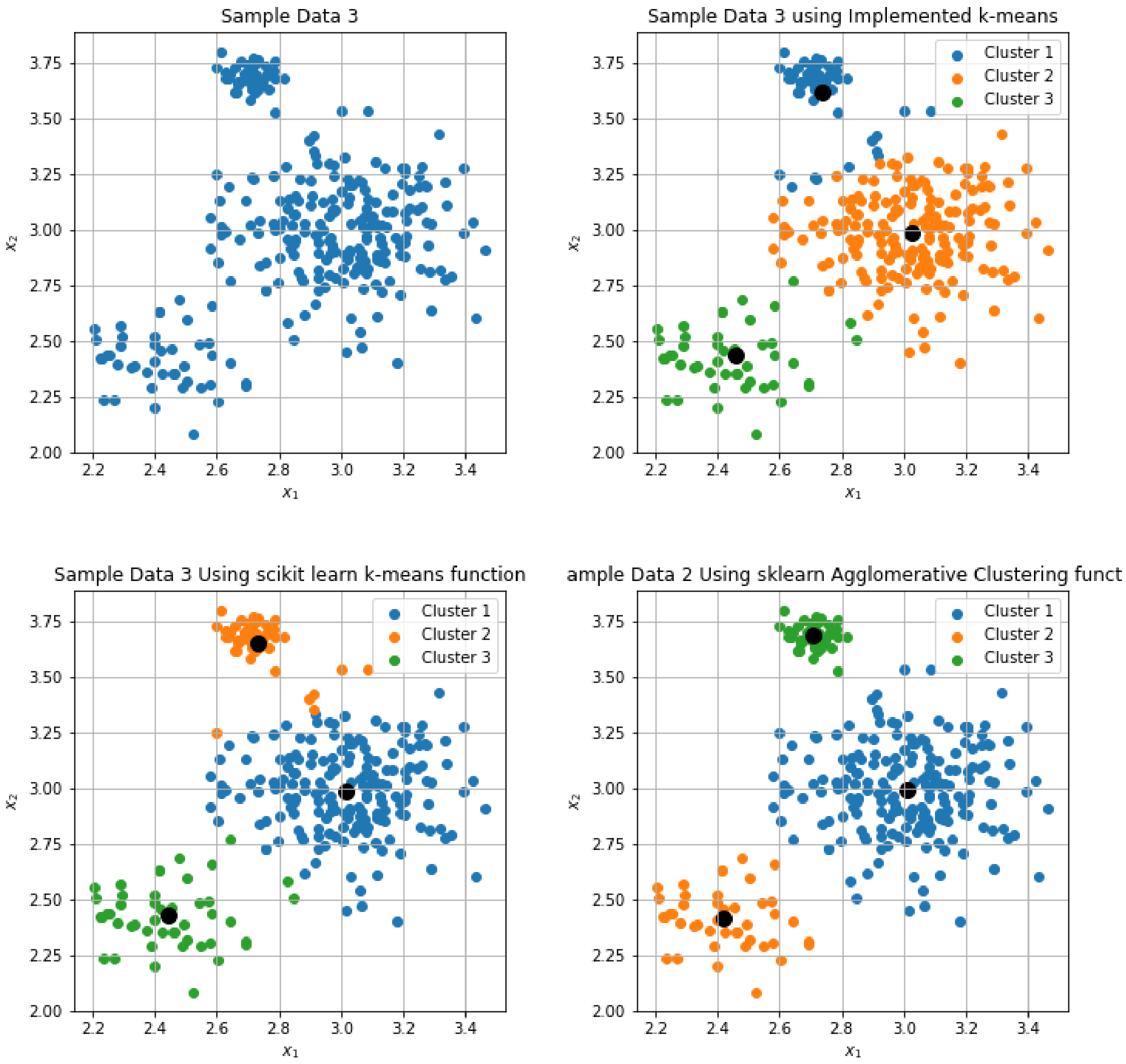

Figure 6: Sample Data 3, Clustered Data using implemented k-means, Scikit Learn k-means, Agglomerative Clustering with 3 clusters

The result of the all three methods are nearly the same and they just are different in just some samples which is due to randomization of the clustering algorithm. All of the methods have clustered our data very well!

** - Important tip

Initial centroids are created randomly in our map, by running all of the algorithms several times, we can see the result is the same as before, just may some few samples change their cluster to their neighbor cluster in our implemented k-means. In scikit learn k-means and Agglomerative clustering which are internal functions of the python, the result is exactly the same in each run.