KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY (KIIT)
Deemed to be University U/S 3 of UGC Act, 1956

Name : Abhishek Mallick

Roll : 21051706

ML_CS-3

Machine Learning

Assignment 3 – K-means Clustering

Github

AIM: The aim of this assignment is to perform K-means clustering on a given dataset.

- Implement the K-means clustering algorithm from scratch using Python, without relying on pre-existing libraries like scikit-learn, TensorFlow, or PyTorch.
- Apply the K-means algorithm to the provided dataset, considering two different values of K: K=2 and K=3.
- Visualize and analyze the clustering results by plotting the dataset points, with different colors representing the respective clusters for each value of K.
- Evaluate the effectiveness of the K-means clustering algorithm on the given dataset by analyzing the obtained clusters.

Dataset used:

https://drive.google.com/file/d/1izjhuVeIVVCIIjWakz1D42bsTBr0CQEt/view?usp=sharing

1. You need to run the k-means algorithm on the dataset for k=2, k=3.

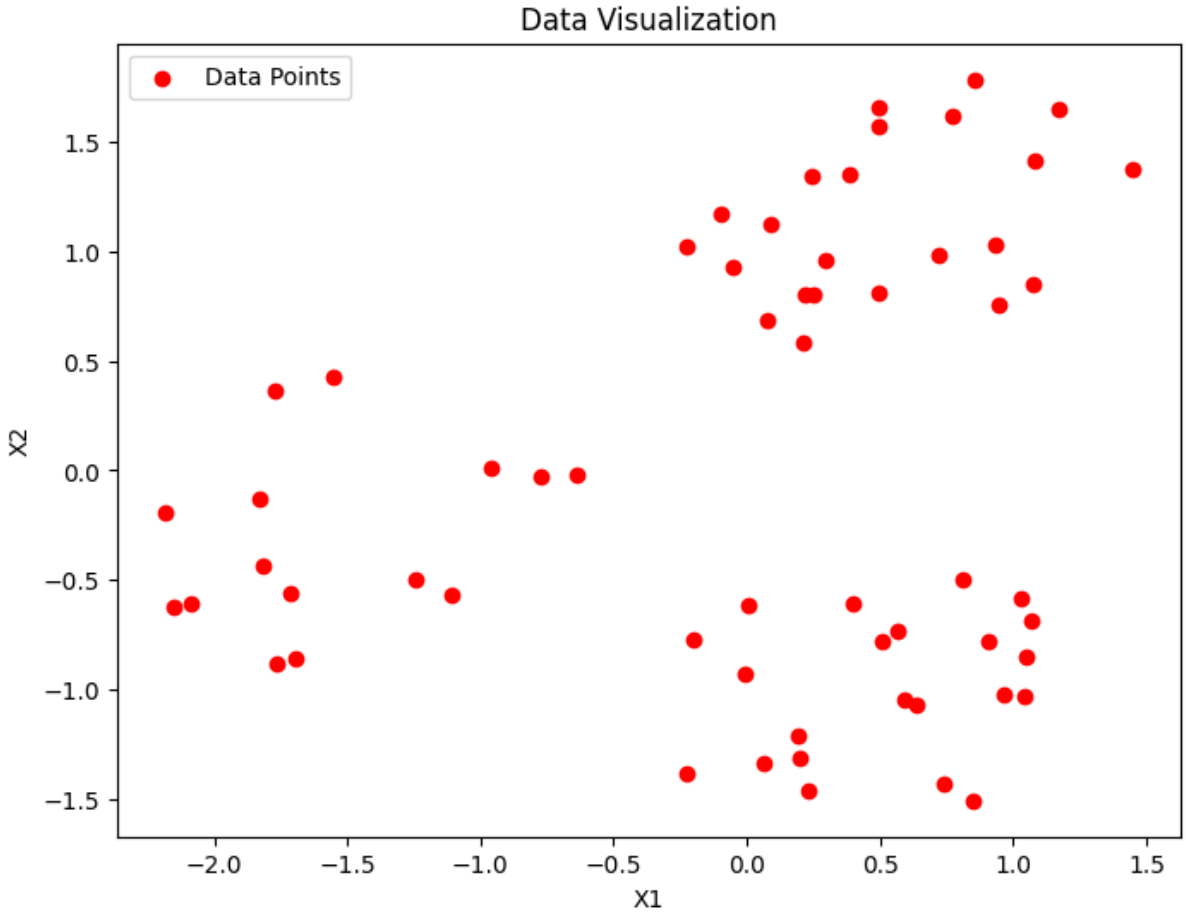
k-means algo representation:

```
def kmeans(data, k, max_iterations=100):
    centroids = initialize_centroids(data, k)

    for _ in range(max_iterations):
        clusters = assign_clusters(data, centroids)
        new_centroids = update_centroids(data, clusters, k)
        if np.all(centroids == new_centroids):
            break
        centroids = new_centroids

    return centroids, clusters
```

Data Vizualization:

```
1 data.head()
```

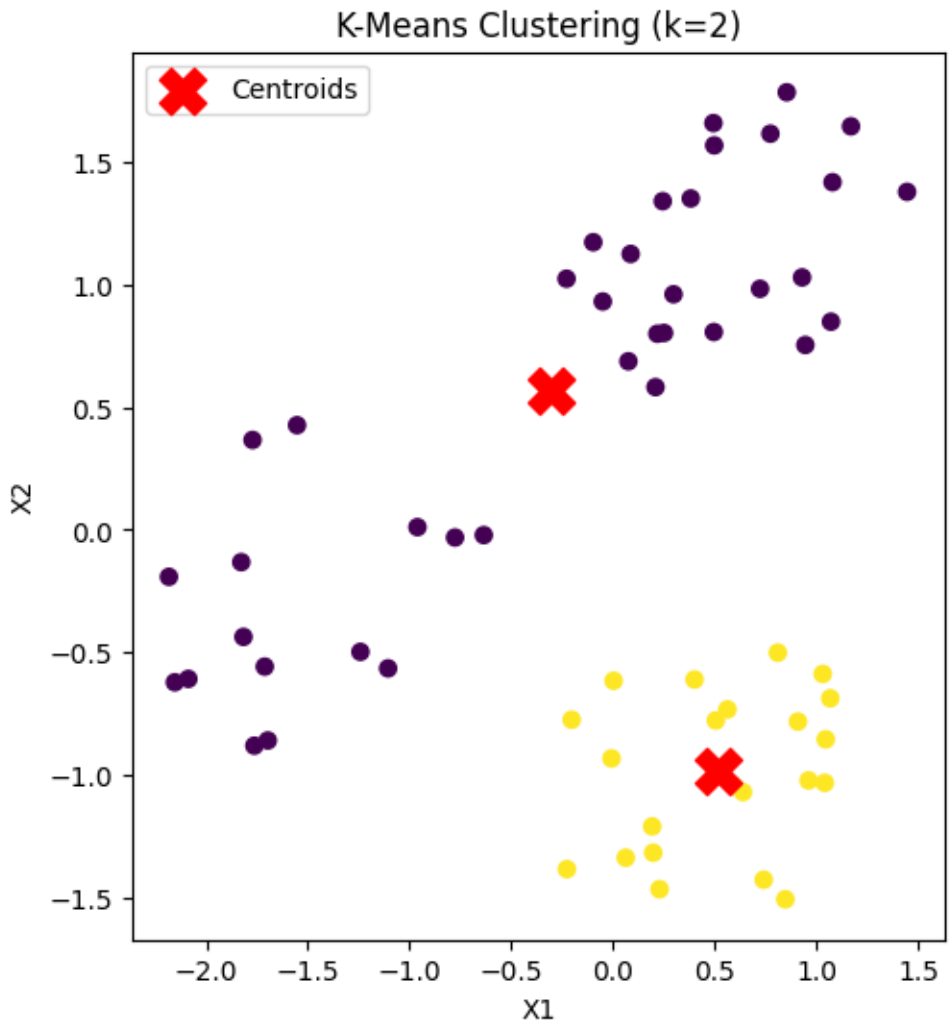
	x1	x2
0	24.412	32.932
1	35.190	12.189
2	26.288	41.718
3	0.376	15.506
4	26.116	3.963



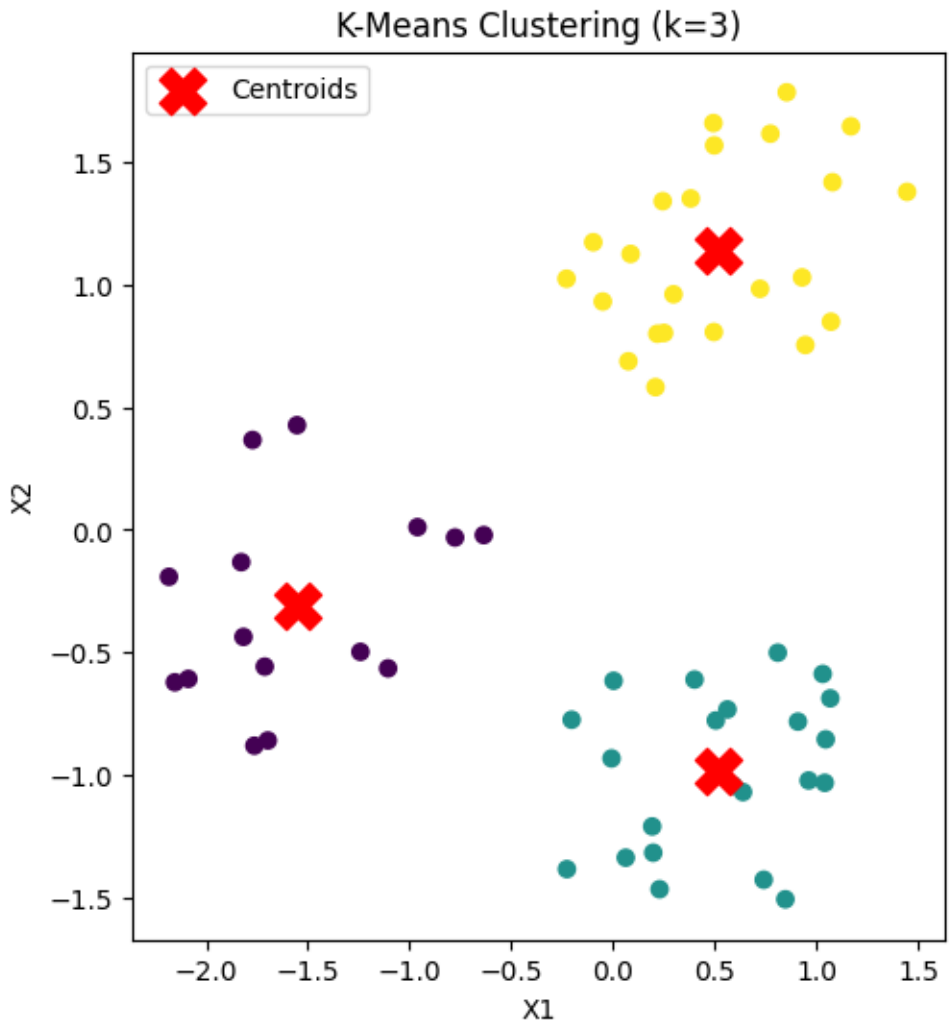
2. You need to run the k-means algorithm on the dataset for k=2, k=3.

```
In [21]:
# Running k-means for cluster 2 & 3
centroids_2, clusters_2 = kmeans(data_normalized.values, 2)
centroids_3, clusters_3 = kmeans(data_normalized.values, 3)
```

K-means Clustering using k=2



K-means Clustering using k=3



Submitted By:

Abhishek Mallick

21051706

ML_CS-3

https://github.com/Abhishek-Mallick/Machine-Learning-KIIT