# K-Means Clustering Analysis

## 1) Problem Statement

## Write a program to do following:

## We have given a collection of 8 points. P1=[0.1,0.6] P2=[0.15,0.71] P3=[0.08,0.9] P4=[0.16, 0.85] P5=[0.2,0.3] P6=[0.25,0.5] P7=[0.24,0.1] P8=[0.3,0.2]. Perform the k-mean clustering with initial centroids as m1=P1=Cluster#1=C1 and m2=P8=cluster#2=C2.

## Answer the following:

## a) Which cluster does P6 belong to?

## b) What is the population of a cluster around m2?

## c) What is the updated value of m1 and m2?

## 2) Libraries Used

Python:  
NumPy: For numerical operations and array manipulation.  
Scikit-learn: For applying the k-means clustering algorithm.

## 3) Theory

K-means clustering is a type of unsupervised learning, which is used when we have unlabeled data. The goal is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

## 4) Methods

The k-means clustering algorithm was applied with two initial centroids defined as m1 (P1) and m2 (P8). Points were assigned to the nearest centroid, and the centroids were updated as the mean of the points in each cluster.

## 5) Advantages

• Simple and efficient for large data sets.  
• Easy to implement and understand.

## 6) Disadvantages

• Sensitive to the selection of initial centroids.  
• May converge to a local optimum depending on initial centroid placement.

## 7) Working

Using the sklearn library, k-means clustering was performed on the dataset with predefined initial centroids. The algorithm iteratively updated the centroids based on the mean of assigned points until convergence.

## 8) Conclusion

The k-means clustering algorithm successfully assigned points to clusters based on their proximity to initial centroids. P6 was assigned to Cluster #1, the population around m2 was 3, and the centroids were updated to new values based on point assignments.