Problem Statement:-

4.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 belongs to? b) What is the population of cluster around m2? c) What is updated value of m1 and m2?

Library:

NumPy.

Theory:

What is Clustering ?

The task of grouping data points based on their similarity with each other is called Clustering or Cluster Analysis. This method is defined under the branch of Unsupervised Learning, which aims at gaining insights from unlabelled data points, that is, unlike supervised learning we don't have a target variable.

Clustering aims at forming groups of homogeneous data points from a heterogeneous dataset. It evaluates the similarity based on a metric like Euclidean distance, Cosine similarity, Manhattan distance, etc. and then group the points with highest similarity score together.

Uses of Clustering

Now before we begin with types of clustering algorithms, we will go through the use cases of Clustering algorithms. Clustering algorithms are majorly used for:

- Market Segmentation – Businesses use clustering to group their customers and use targeted advertisements to attract more audience.

- Market Basket Analysis – Shop owners analyze their sales and figure out which items are majorly bought together by the customers. For example, In USA, according to a study diapers and beers were usually bought together by fathers.

- Social Network Analysis – Social media sites use your data to understand your browsing behaviour and provide you with targeted friend recommendations or content recommendations.

- Medical Imaging – Doctors use Clustering to find out diseased areas in diagnostic images like X-rays.

Diagram:
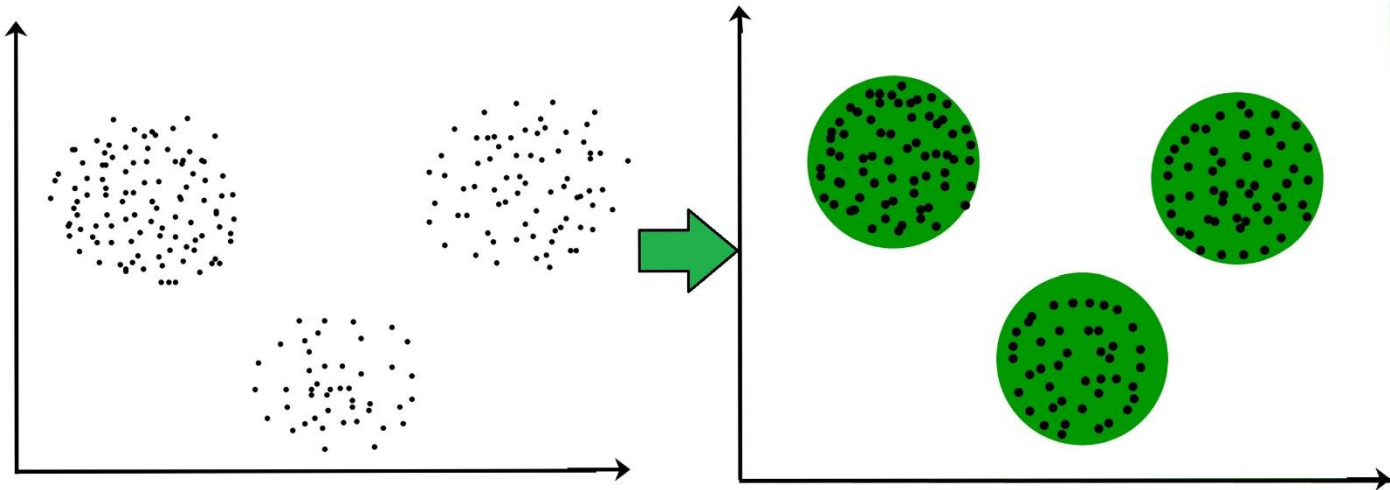
Conclusion:

Clustering is a powerful unsupervised machine learning technique used to organization unlabelled data factors into similar clusters. The clustering results, together with the temporal relations of the shots, are used to build the scene transition graph. Each node represents a collection of shots while an edge reflects the flow of story from one node to the next. Data scientists use clustering algorithms to gain valuable insights and make informed decisions to solve real-world problems.