

MACHINE LEARNING & IMAGE PROCESSING
WEEK 06 - PART III

K MEANS CLUSTERING

Thakshila Dasun
BSc. Hons in Mechanical Engineering
(Mechatronics Specialization)
CIMA, UK
Academy of Innovative Education



Introduction to Clustering

- A cluster refers to a small group of objects.
- Clustering is grouping those objects into clusters.
- In order to learn clustering, it is important to understand the scenarios that lead to cluster different objects.
- Clustering is dividing data points into homogeneous classes or clusters:
 - Points in the same group are as similar as possible
 - Points in different group are as dissimilar as possible

Application of Clustering

- Clustering is used in almost all the fields. You can infer some ideas from Example 1 to come up with lot of clustering applications that you would have come across.
 1. Clustering helps marketers improve their customer base and work on the target areas. It helps group people (according to different criteria's such as willingness, purchasing power etc.) based on their similarity in many ways related to the product under consideration.
 2. Clustering helps in identification of groups of houses on the basis of their value, type and geographical locations.
 3. Clustering is used to study earth-quake. Based on the areas hit by an earthquake in a region, clustering can help analyse the next probable location where earthquake can occur.

Clustering Algorithms

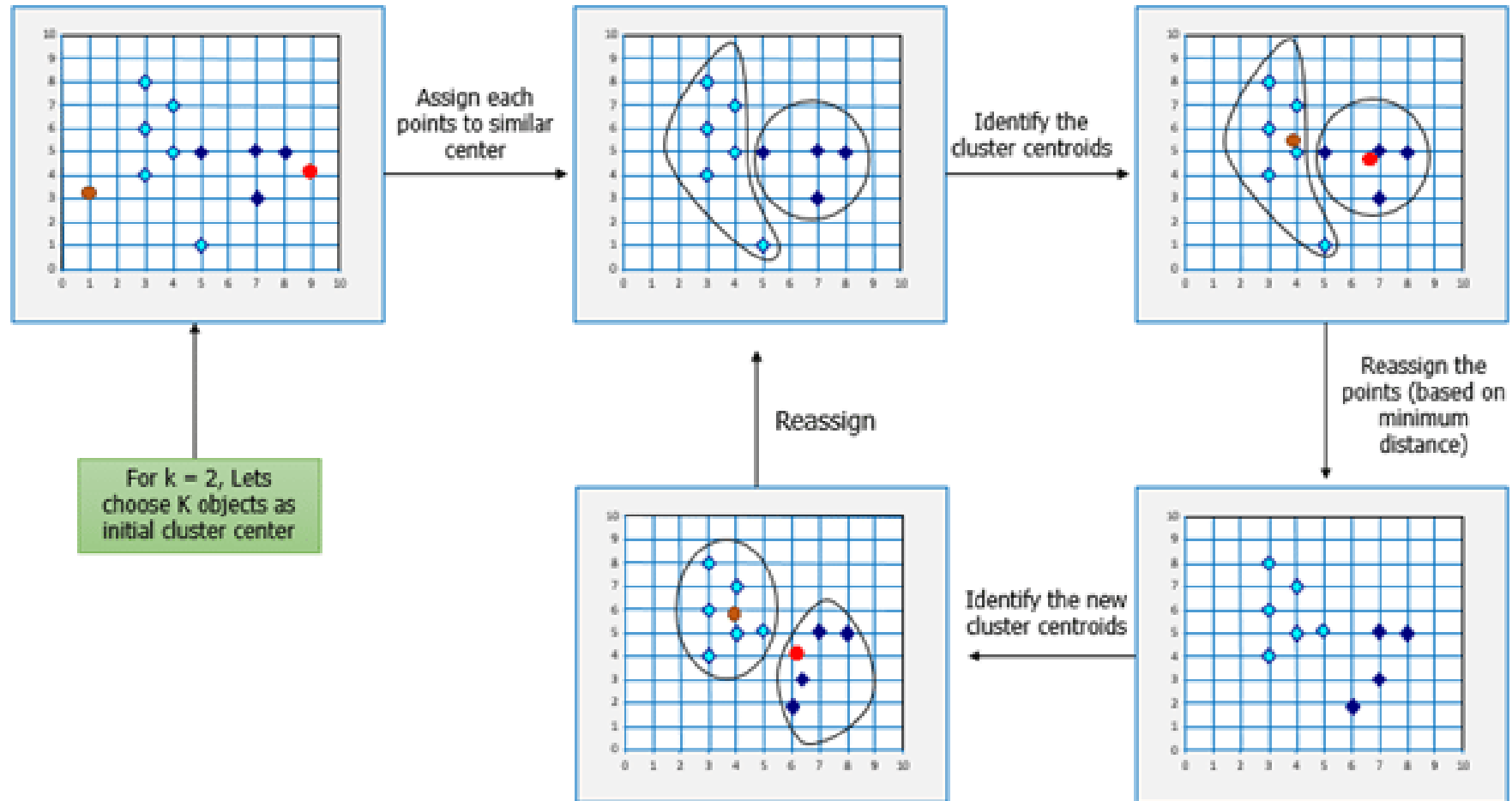
- A Clustering Algorithm tries to analyze natural groups of data on the basis of some similarity.
- It locates the **centroid** of the group of data points.
- To carry out effective clustering, the algorithm evaluates the distance between each point from the centroid of the cluster.
- The goal of clustering is to determine the intrinsic grouping in a set of **unlabeled** data.



K-means Clustering (1)

- If k is given, the K-means algorithm can be executed in the following steps:
 1. Partition of objects into k non-empty subsets
 2. Identifying the cluster centroids (mean point) of the current partition.
 3. Assigning each point to a specific cluster
 4. Compute the distances from each point and allot points to the cluster where the distance from the centroid is minimum.
 5. After re-allotting the points, find the centroid of the new cluster formed.

K-means Clustering (2)



Let's Try this example



How to find the Optimal K Value

- For any k clusters, the value of k should be such that even if we increase the value of k from after several levels of clustering the **Variance** remains constant. The achieved point is called the **“Elbow”**.

