

# Lab Assignment.

Date \_\_\_\_\_  
Page \_\_\_\_\_

K-means clustering algorithm.

⇒ K-means clustering algorithm is a popular method of grouping data by assigning observations to clusters based on proximity to the cluster's center. It is an unsupervised learning technique that groups data points into clusters based on similarity.

The main objective of k-means algorithm is to minimize the sum of distances between the points and their respective cluster centroid.

Optimization plays a crucial role in k-means clustering algorithm. The goal of optimization process is to find the best of centroids that minimizes the sum of squared distances between each datapoint and its closest centroid.

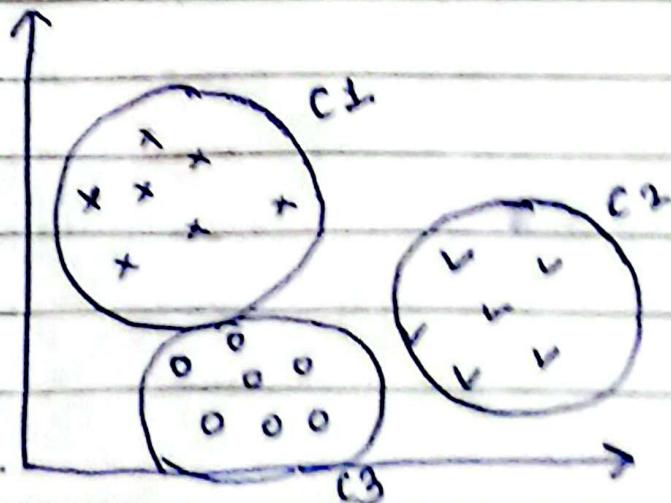


Fig: K-means clustering through graph.

Code:

- i) Import matplotlib.pyplot as plt  
~~X = [4, 8, 16, 20] [4, 2, 1, 2, 10, 9, 10, 9, 5, 6, 6]~~  
~~Y = [2, 8, 9, 10] [2, 3, 4, 1, 2, 3, 4, 1, 9, 9, 7, 9]~~  
// Sample data  
plt.scatter(x, y) // visualize data
- ii) from sklearn.cluster import kmeans  
  
data = list(zip(x, y))  
inertias = [] // elbow method  
  
for i in range(1, 11):  
 kmeans = kmeans(n\_clusters=i)  
 kmeans.fit(data)  
 inertias.append(kmeans.inertia\_)

plt.plot(range(1, 11), inertias, marker='o')
plt.title('k-means')
plt.xlabel('number of clusters')
plt.show()
- iii) kmeans = kmeans(n\_clusters=2)
kmeans.fit(data)

plt.scatter(x, y, c=kmeans.labels)
plt.show

## Explanation.

Date \_\_\_\_\_  
Page \_\_\_\_\_

- 1) Imported libraries which are required
- 2) prepared the datasets randomly in terms of and  $(x, y)$
- 3) visualizing using plot.
- 4) Finding best number of clusters using elbow method.  
This method provides best clusters, also removes under and over fitting,
- 5) plotting the graph.
- 6) choosing the best  $k$  (based on elbow) and train k-means
- 7) visualizing and print clusters.

$$\text{Inertia} = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$