Roll No 41163

Page No.
Date: / /

Assignment A4

Execution Date: 17/03/2021
Submission Date: 21/05/2021

Title: K-means algorithm for clustering

Problem Statement:
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 K-means clustering with initial centroids as m1 = P1 = cluster#1 and m2 = P8 = cluster#2

Answer the following:
1) What cluster does P6 belong to?
2) What is the population of cluster around m2?
3) What is the updated value of m1 and m2?

Objective:
To understand and implement K-means algorithm.

Outcome:
Students will learn how K-means clustering algorithm works.

Page No.
Date: / /

Software Requirements:
- Jupyter notebook
- Python 3.8.5
- 64 bit operating system

Hardware Requirements:
- Computer with 64 bit processor

Theory:

K-means clustering is one of the easimplest and popular unsupervised machine learning algorithm.

A cluster refers to the group of data points aggregated together because of certain similarities.

K-means clustering algorithm identifies K-number of centroids and then allocate every data point to the nearest cluster while keeping the centroids as small as possible.

To process the learning data, the K-means algorithm in data mining starts with the first group of randomly selected centroids, which are used as the beginning points for every cluster and then performs iterative calculations to optimize the position of centroids.

It halts creating and optimizing clusters when:

- The centroids have stabilized
- The defined number of iterations have been achieved.

Page No.
Date: / /

Algorithm:

1) Choose K-objects from data as initial clusters centers.

2) Assign each object to a closest cluster

3) Calculate mean of all objects in cluster and update centroid,

4) Repeat.

Example:

Input Data = Data in problem statement

Distances

Data point	Distance C1	Distance C2
(0.1, 0.6)	0	0.45
(0.15, 0.71)	0.12	0.53
(0.08, 0.9)	0.30	0.73
(0.16, 0.85)	0.25	0.67
(0.2, 0.3)	0.32	0.14
(0.25, 0.5)	0.18	0.30
(0.24, 0.1)	0.51	0.12
(0.3, 0.2)	0.45	0

Assign clusters:

Cluster 1:

(0.1, 0.6)	(0.08, 0.9)	(0.25, 0.5)
(0.15, 0.71)	(0.16, 0.85)	

Page No.
Date: / /

Cluster 2

(0.2, 0.3) (0.24, 0.1) (0.3, 0.2)

Calculate new values of centroids

$m_1 = (0.148, 0.712)$

$m_2 = (0.24667, 0.20)$

Test Cases:

Description	Expected O/P	Actual O/P
Perform clustering with given data	Divide data into 2 clusters	Successful

~~Perform cl~~

Conclusion:

We have studied the K-means clustering algorithm and implemented it.