GeeksforGeeks
A computer science portal for geeks

ML | K-Medoids clustering with example

K-Medoids (also called as Partitioning Around Medoid) algorithm was proposed in 1987 by Kaufman and Rousseeuw. A medoid c̣ whose dissimilarities with all the other points in the cluster is minimum.

The dissimilarity of the $medoid(Ci)$ and $object(Pi)$ is calculated by using $E = |Pi - Ci|$

Hire with us!

The cost in K-Medoids algorithm is given as −

$$c = \sum_{Ci} \sum_{Pi \in Ci} |Pi - Ci|$$

.

Algorithm:

1. Initialize: select k random points out of the n data points as the medoids.

2. Associate each data point to the closest medoid by using any common distance metric methods.

3. While the cost decreases:

For each medoid m, for each data o point which is not a medoid:

1. Swap m and o, associate each data point to the closest medoid, recompute the cost.

2. If the total cost is more than that in the previous step, undo the swap.

	X	Y	Dissimilarity From C1	Dissimilarity From C2
0	7	6	6	2
1	2	6	3	7
2	3	8	4	8
3	8	5	6	2
4	7	4	4	0
5	4	7	4	6
6	6	2	5	3
7	7	3	5	1
8	6	4	3	1
9	3	4	0	4

Each point is assigned to the cluster of that medoid whose dissimilarity is less.

The points 1, 2, 5 go to cluster C1 and 0, 3, 6, 7, 8 go to cluster C2.

The cost C = (3 + 4 + 4) + (3 + 1 + 1 + 2 + 2)

C = 20

Step #3: Now randomly select one non-medoid point and recalculate the cost.

Let the randomly selected point be (7, 3). The dissimilarity of each non-medoid point with the medoids

− c1 (3, 4) and c2 (7, 3) is calculated and tabulated.

	X	Y	Dissimilarity From C1	Dissimilarity From C2
0	7	6	6	3
1	2	6	3	8
2	3	8	4	9
3	8	5	6	3
4	7	4	4	1
5	4	7	4	7
6	6	2	5	2
7	7	3	-	-
8	6	4	3	2
9	3	4	-	-

Each point is assigned to that cluster whose dissimilarity is less. So, the points 1, 2, 5 go to cluster C1 and 0, 3, 6, 7, 8 go to cluster C2.
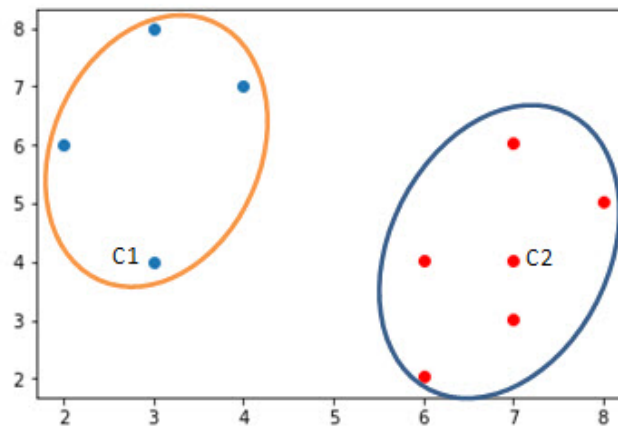
The cost C = (3 + 4 + 4) + (2 + 2 + 1 + 3 + 3)

C = 22

Swap Cost = Present Cost − Previous Cost

= 22 − 20 = **2 >0**

As the swap cost is not less than zero, we undo the swap. Hence (3, 4) and (7, 4) are the final medoids. The clustering would be in the following way

The **time complexity** is $O(k * (n - k)^2)$.

Advantages:

1. It is simple to understand and easy to implement.
2. K-Medoid Algorithm is fast and converges in a fixed number of steps.
3. PAM is less sensitive to outliers than other partitioning algorithms.

Disadvantages:

1. The main disadvantage of K-Medoid algorithms is that it is not suitable for clustering non-spherical (arbitrary shaped) groups of objects. This is because it relies on minimizing the distances between the non-medoid objects and the medoid (the cluster center) – briefly, it uses compactness as clustering criteria instead of connectivity.
2. It may obtain different results for different runs on the same dataset because the first k medoids are chosen randomly.

Recommended Posts:

K means Clustering - Introduction

Clustering in R Programming

Analysis of test data using K-Means Clustering in Python

Clustering in Machine Learning

Different Types of Clustering Algorithm

ML | Unsupervised Face Clustering Pipeline

ML | Determine the optimal value of K in K-Means Clustering

DBSCAN Clustering in ML | Density based clustering

ML | Hierarchical clustering (Agglomerative and Divisive clustering)

ML | Mini Batch K-means clustering algorithm

Image compression using K-means clustering

ML | Mean-Shift Clustering

Implementing Agglomerative Clustering using Sklearn

ML | OPTICS Clustering Implementing using Sklearn

ML | OPTICS Clustering Explanation

ML | V-Measure for Evaluating Clustering Performance

ML | Spectral Clustering

ML | Types of Linkages in Clustering

ML | Fuzzy Clustering

ML | Classification vs Clustering

kanakalathav99
Check out this Author's contributed articles.

If you like GeeksforGeeks and would like to contribute, you can also write an article using contribute.geeksforgeeks.org or mail your article to contribute@geeksforgeeks.org. See your article appearing on the GeeksforGeeks main page and help other Geeks.

Please Improve this article if you find anything incorrect by clicking on the "Improve Article" button below.

Article Tags : Machine Learning

Practice Tags : Machine Learning

1

1

Based on **1** vote(s)

☐ To-do ☐ Done

Feedback/ Suggest Improvement   Improve Article

Please write to us at contribute@geeksforgeeks.org to report any issue with the above content.

Writing code in comment? Please use ide.geeksforgeeks.org, generate link and share the link here.

Load Comments

A computer science portal for geeks

5th Floor, A-118,
Sector-136, Noida, Uttar Pradesh - 201305
feedback@geeksforgeeks.org

COMPANY

About Us
Careers
Privacy Policy
Contact Us

LEARN

Algorithms
Data Structures
Languages
CS Subjects
Video Tutorials

PRACTICE

Courses
Company-wise
Topic-wise
How to begin?

CONTRIBUTE

Write an Article
Write Interview Experience
Internships
Videos

A computer science portal for geeks

Let's consider the following example:

	X	Y
0	7	6
1	2	6
2	3	8
3	8	5
4	7	4
5	4	7
6	6	2
7	7	3
8	6	4
9	3	4

If a graph is drawn using the above data points, we obtain the following:
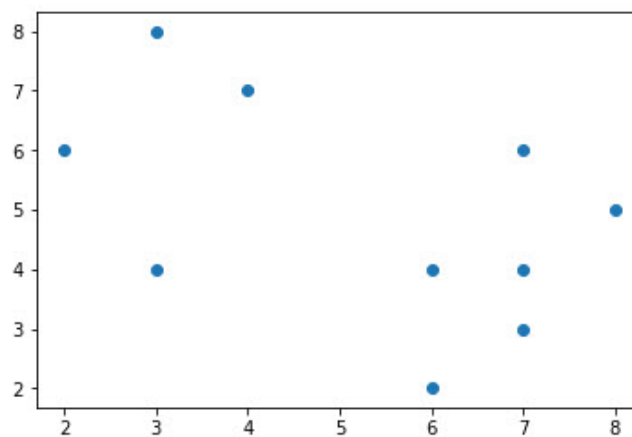
Step #1: k = 2

Let the randomly selected 2 medoids be `c1` -(3, 4) and `c2` -(7, 4).

Step #2: Calculating cost.

The dissimilarity of each non-medoid point with the medoids is calculated and tabulated: