

K-Means Algorithms Unsupervised Learning +4

## What is the k-Means algorithm and how does it work?

Answer

Follow · 48

Request

1



### 20 Answers

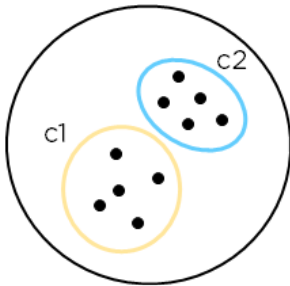


Anirudh Sharma, Exploring Data Science  
Answered May 4, 2018



**K-Means** is one of the simplest unsupervised learning **algorithms** that solve the well known **clustering** problem.

- The procedure follows a simple and easy way to classify a given data set through a certain number of clusters.



### K Means Algorithm:

#### STEP 1:

We randomly pick '**K**' cluster centers (centroids).

Let's assume these are  $c_1, c_2, \dots, c_k$  and we can say that;

$$C = c_1, c_2, \dots, c_k$$

'C' is the set of all centroids.

#### STEP 2:

In this step, we assign each data point to closest center, this is done by calculating Euclidean distance.

$$\arg \min_{c_i \in C} \text{dist}(c_i, x)^2$$

Where  $\text{dist}()$  is the Euclidean distance.

#### STEP 3:

In this step, we find the new centroid by taking the average of all the points assigned to that cluster.

### Related Questions

[How is the k-nearest neighbor algorithm different from k-means clustering?](#)

[What is a k-means algorithm?](#)

[Can someone explain k-means++ algorithm?](#)

[What is the difference between k-means clustering and parallel k-means clustering algorithm?](#)

[How are k-means clustering algorithms sensitive to outliers?](#)

[What is K algorithm?](#)

[How does the K- Nearest Neighbour algorithm work?](#)

[What is the mean-shift algorithm?](#)

[Which is the best adaptive k-means clustering algorithm \(that k is automatically chosen\)?](#)

[What are the principles of Spherical K-means algorithm?](#)

[+ Ask New Question](#)

[More Related Questions](#)

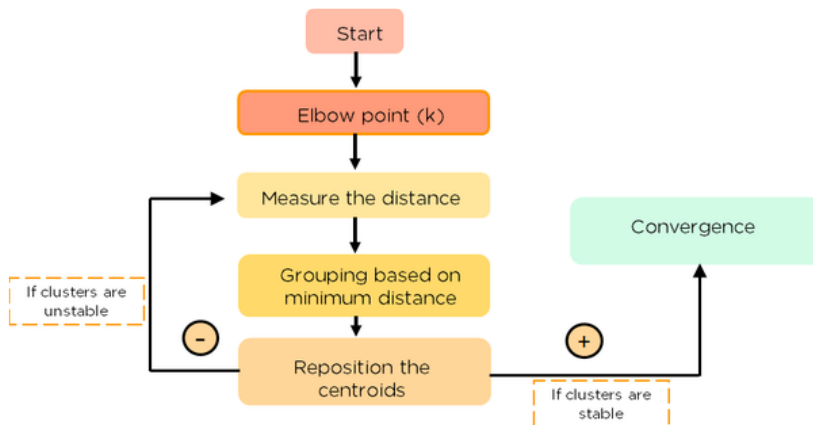
'Si' is the set of all points assigned to the ith cluster.

#### STEP 4:

In this step, we repeat step 2 and 3 until none of the cluster assignments change

That means until our clusters remain stable, we repeat the algorithm

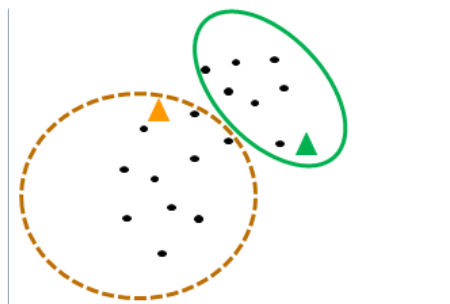
#### HOW DOES K-MEANS WORK?



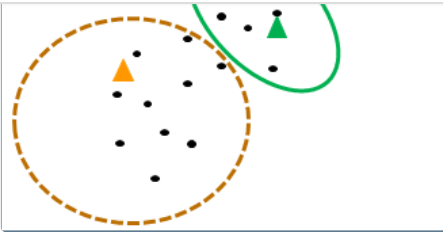
**STEP 1:** First, we need to randomly initialize two points called the cluster centroids.



**STEP 2:** Now, based upon the distance from the orange cluster centroid or green cluster centroid, it will group itself into that particular group.

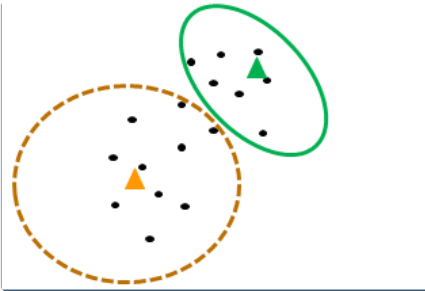


**STEP 3: Move Centroids** - Now, you will take the two cluster centroids and iteratively reposition them for optimization.



**STEP 4:** Repeat previous two steps iteratively till the cluster centroids stop changing their positions and become static.

**STEP 5:** Once the clusters become static then k-means clustering algorithm is said to be converged.



Hope you understand the process.

To learn more about K-Means clustering watch the video given below.

If you are interested to learn more about Machine Learning and gain an in-depth knowledge, check out our [Machine Learning Certification Training](#) .

1.8k views · View Upvoters

Upvote · 17 Share



Add a comment...

Recommended All



Bhanupriya Katrapally, Datascience Trainer at ExcelR Solutions  
Answered Apr 18



**K-means** is one of the clustering methods. Clustering is an unsupervised machine learning algorithm. Clustering is nothing but grouping similar records together in a given dataset.

Broadly clustering algorithms are divided into Hierarchical and Non-Hierarchical algorithms.

**Hierarchical** method is useful when you want to arrange the clusters into a natural hierarchy, based on the distance calculation between observations in a

Home<sup>1</sup>

Answer



Spaces



Notifications

Search Quora



Add Question or Link

when you want to divide the dataset into a pre-specified number of clusters.

**K-Means is a non-hierarchical clustering algorithm.** It is an iterative algorithm which tries to partition the dataset into a pre-defined number of non-overlapping clusters(k) by minimizing the sum of squared distances from each data point to the cluster centroid(Within sum of squared distances- WSS).

#### How K means Works?

1. Pre-define the no. of clusters (k)
2. Initialize cluster centroids by randomly selecting K data points.
3. Calculate the WSS distances between each data point to all the cluster centroids.
4. Assign each data point to the closest centers

Steps 3 and 4 will be iterated until there is no reassignment of data points is required.

#### Key Points

- Since this algorithm is based on distance measure, it is recommended to **standardize the data**.
- As K-means is iterative in nature and random initialization of centroids at the start of the algorithm, different initialization may lead to different clusters. Therefore it is recommended to run the algorithm several times and pick up the run which is giving the minimum WSS.

**The K-Mean Clustering Algorithm is explained clearly in this video below.**

**To learn more about K - means algorithm and gain expertise in it, Subscribe to our [YouTube Channel](#) and check out our ExcelR's [Data Science course](#)**

354 views · View Upvoters

Upvote · 6 Share



Add a comment...

Recommended All



Annalyn Ng, Data Scientist  
Updated Aug 2, 2016



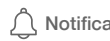
(Tutorial entry taken from: [Annalyzing Life / Data Analytics Tutorials & Experiments for Layman](#) )

Home<sup>1</sup>

Answer



Spaces



Notifications

Search Quora



Add Question or Link

might be interested in. However, doing this for each individual is grossly inefficient. A better solution would be to categorize his customers into groups, with each group having similar preferences. This would allow him to reach out to more customers with each product recommendation.

The problem is, the storekeeper does not know **1) how his customer...** [\(more\)](#)

Upvote · 70 Share



Add a comment...

Recommended All



Shehroz Khan, ML Researcher, Postdoc @U of Toronto  
Answered May 6, 2015



Non-mathematical answer.

Imagine you have some data (2D for simplicity), you don't know the labels but you have some knowledge about how many clusters or structures it may have (K). Now you want to construct K clusters so as to group the data into them using some notion of similarity. You are free to choose the similarity metric, euclidean, mahalanobis, cosine or whatever, but each similarity metric has its own intricacy. K-means construct spherical clusters (circle in 2D) so you know what kind of clusters you may get.

Now you want to start, either choose K random initial points or choose ... [\(more\)](#)

Upvote · 11 Share



Add a comment...

Recommended All



Avril Kenney, studied at Massachusetts Institute of Technology  
Answered Apr 19, 2014



The k-means algorithm is a **clustering algorithm**. That means that you have a bunch of points in some space, and you want to guess what groups they seem to be in. For example, say we have these points:

```

1   o
2  o oo
3   o o
4
5           oo
6         oo  o

```

As a human, you can easily look at those and say that the ones in the top left are a cluster and the ones in the bottom right are a cluster, but if there were lots more clusters, or if they overlapped, or if they were in a 3-dimensional or much higher dimensional space, it would be harder.... [\(more\)](#)

Upvote · 10 Share



Add a comment...

Recommended All



Rishiraj Surti  
Answered Jul 13, 2015







The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori.

The main idea is to define k centroids, one for each cluster.

These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to... [\(more\)](#)

 Upvote · 7     Share      



Recommended All

Top Stories from Your Feed