



What Are The Main Difference Between K Means And KNN?



K MEANS VS KNN

Difference between K Means
and KNN and how they
works

June 9, 2021 | Machine learning



K Means and KNN are both machine learning algorithms. One is an Unsupervised algorithm used for clustering while the other is a supervised algorithm used for classification. But because of their name similarities, many newbies get confused. So, in this article we will see the complete difference between K means and Knn machine learning, also we will see how they work. Hope you will like it 😊

What Is K-Means Clustering In Machine Learning?

K-Means is a clustering algorithm, which is a part of unsupervised machine learning(data with no labels). It is used to create clusters out of lots of unlabeled data points where data holding a similar property or pattern are stored in the same cluster. "K" in K means represents the number of clusters it forms, which is decided by the practitioner.

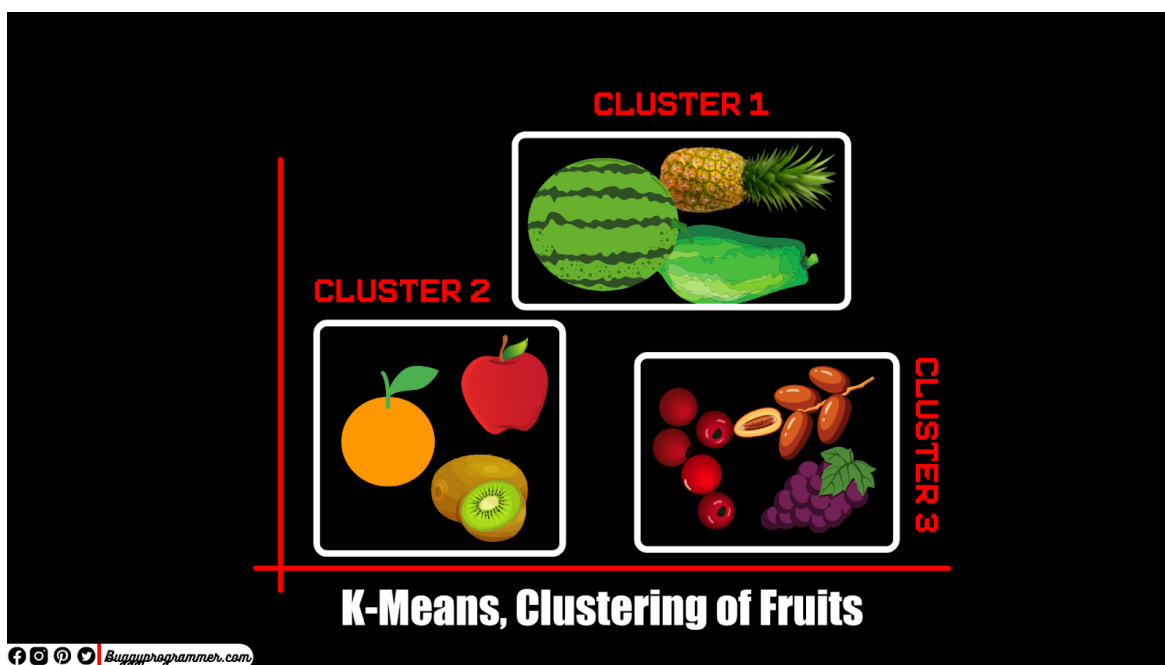
For example, if you have data of lots of fruit without labels(means you just have their property like shape, size, color, taste, etc and you don't know which fruit it is) then based on their property it will create different clusters and put the fruits with same property in one cluster and others in other clusters.

Avaya

Download The White Paper

Download

Also Read -> [What is Docker and why we need a beginners guide](#)



Clustering in K Means

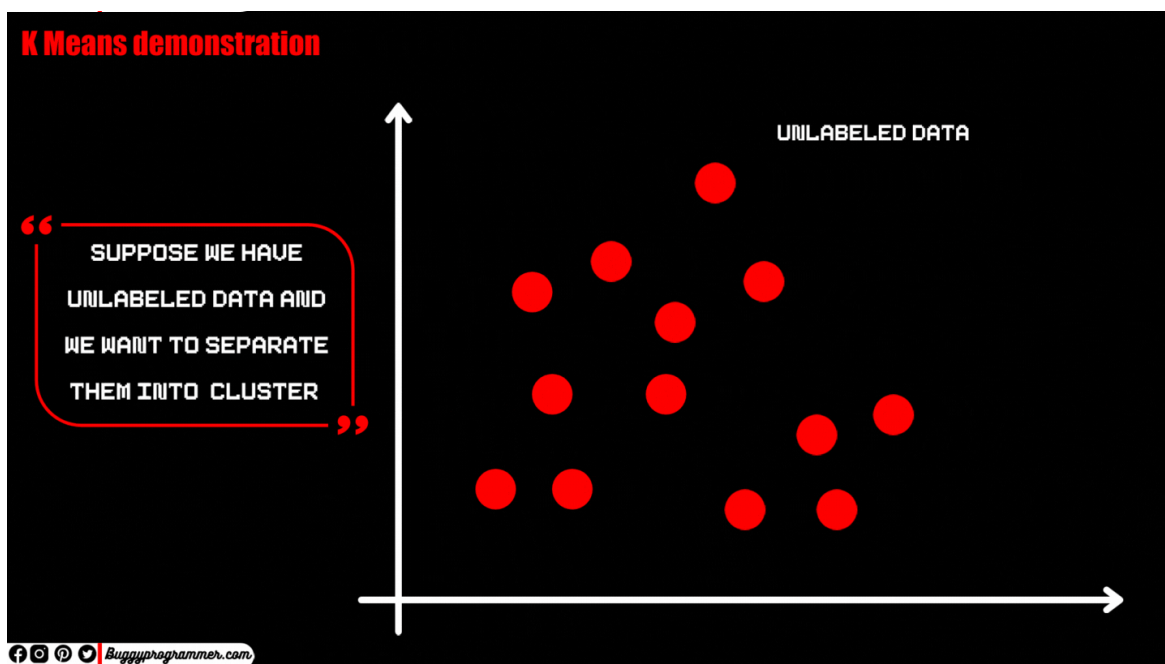
How Does KMeans Work?

It is very easy to understand how K means work. To understand it better, let's see it's process step by step.

Step 1 → After collecting and feeding the data to the model we first decide the value of "K" which is nothing but the number of clusters. The value of "K" is generally decided on the basis of understanding of domain and data. There is also one other method to decide the value of "K" which we will discuss later in this article.

Step 2 → Then we randomly choose the K number of Centroid (center of data points, initially we choose it randomly so it is actually not in a center) which represents the center of a particular cluster. For example, if we choose K's value to be 2 then we will have two clusters and each cluster will have one Centroid.

Step 3 → Then we calculate the distance of each data point from each centroid we created. And then we assign the data points to a certain cluster based on the closest distance it has from a centroid. For example If the distance of a data point 'P' from the centroid "A" is 5 and from the centroid "B" is 7 then the data point 'P' will be assigned to centroid "A".



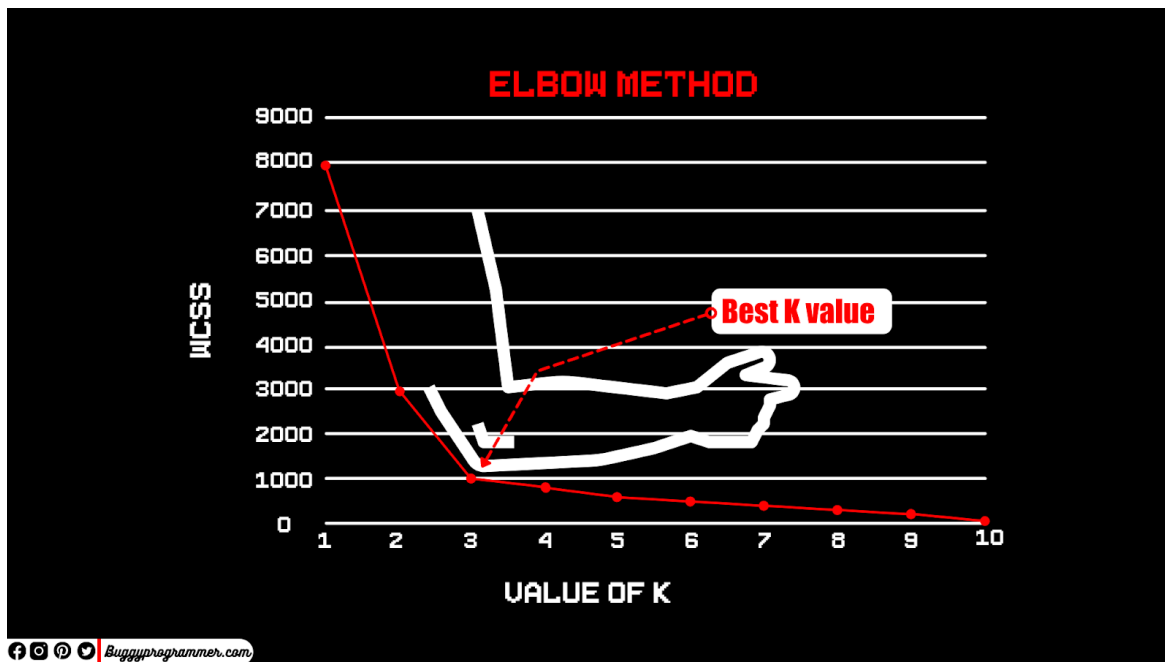
How K Means works

Now we will look at the other method for choosing the best "K" value for your model. The method is called the "Elbow" method, quite a weird name right! XD, we will see why it is.

We decide the best value of K by first calculating **WCSS**(Within Cluster Sum of Square) of a model with different values of "K" .

(WCSS measures the sum of squared average distance of each point from its centroid within a cluster. The smaller the value the more compact it is and vice versa. In short this metric is used to measure the cluster's evaluation, if you want to learn more about it then you can read it [here](#).)

And then we plot the line graph out of it and search for a point where wcss score suddenly stops decreasing, usually this forms a curve like an elbow and that's the reason it's called elbow curve or elbow method.



How Elbow Method works

What Is KNN Classification?

K NN stands for K nearest neighbor, which is a classification algorithm of Supervised Learning. It is used to classify a data point to a certain class, based on its property and patterns, for example, if you have two classes, Dog, and Cat and you passed an image of a cat then it will be classified as a cat class.

Just like K means, "K" in Knn also has a meaning, actually, it is a parameter that refers to the number of nearest neighbors to include in the majority of the voting process.

Also Read → [Deep learning complete guide for beginners without math](#)

How Does KNN Work?

To understand how K NN works let's see its process step by step

Step 1 → First we choose the data point we want to classify, let's name it "M" for a while.

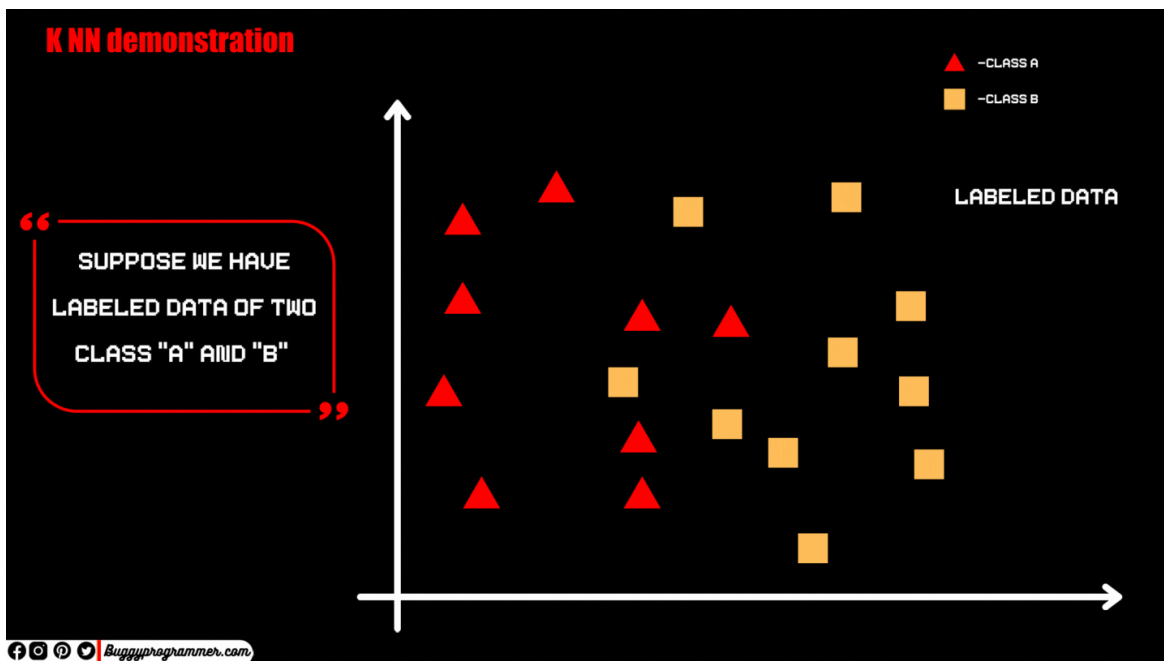
Step 2 → After that just like K Means we then decide the value of K but here it is not the value for cluster, instead it is the number of nearest neighbours to a point we want to classify which in our case is "M". The value of K is decided by the practitioner, depending on the data and domain.



Also it is believed to be good practise to keep the value of "K" as an odd number because we classify any value to a class based on the maximum number of nearest data points of certain class to the point we want to classify and if we choose the value of "K" as an even number then we it could

For example, if you want to classify the point "P" to the class "A" or "B" and the value of "K" is 5 then the model will pick 5 nearest values to the point "A". Those points could be from any or both class, depending on its position

Step 3 → After deciding the value of "K", (which is K nearest data points of available class to the data points we are classifying "M") then let's create an imaginary circle covering those K values around the point "M". And now we will count whose class's data points are maximum inside the circle. The data point will be assigned to the class which has the maximum data point inside that circle.

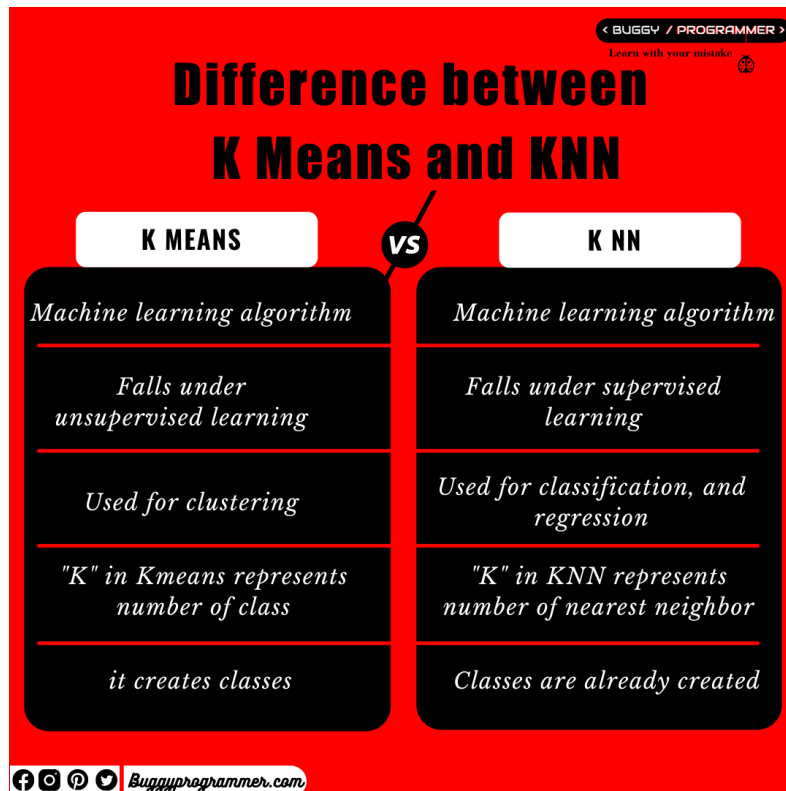


How KNN classification works

Difference Between K Means And KNN?

The big main difference between K means and KNN is that K means is an unsupervised learning clustering algorithm, while KNN is a supervised learning classification algorithm. K means creates classes out of unlabeled data while KNN classifies data to available classes from labeled data.

Also, read → [Difference between Java and Javascript](#)



Difference between K Means and KNN

Download

Avaya

Download 1
White Pape

Download

Download

Share this post



Darkstar



Data Scientist with 3+ years of experience in building data-intensive applications in diverse industries. Proficient in predictive modeling, computer vision, natural language processing, data visualization etc. Aside from being a data scientist, I am also a blogger and photographer.

Read next...

Pandas Join Vs Concat For Aggregating Data In A Simple Way

READ MORE »

July 13, 2022

Namedtuple Vs Dict: All You Need To Know About Data Types In Python

READ MORE »

July 11, 2022

What Is The Conditional Where Clause In Postgresql? Simple Explanation

READ MORE »

July 9, 2022

How To Import Module From Parent Directory In Python In 2022?

READ MORE »

July 8, 2022

Subscribe ▼

Login



B I U        



1 COMMENT



Newest ▾



Is K Means Supervised Or Unsupervised? – Almazrestaurant

7 months ago

[...] The big main difference between K means and KNN is that K means is an unsupervised learning clustering algorithm, while KNN is a supervised learning classification algorithm. K means creates classes out of unlabeled data while KNN classifies data to available classes from labeled data. via [...]



0



Reply

 ezoic

report this ad

KEEP IN TOUCH



About

About
Contact Us
Privacy Policy

Let's Talk

Support Mail
Query Email

Support Me

Pateron
Buy me a coffee

 #Buggy programmer

Never be afraid of making mistakes. It's the one who teaches you the best.

X