



What is KNN Algorithm?

KNN – K Nearest Neighbors, is one of the simplest **Supervised** Machine Learning algorithm mostly used for

Classification



It classifies a data point based on how its neighbors are classified

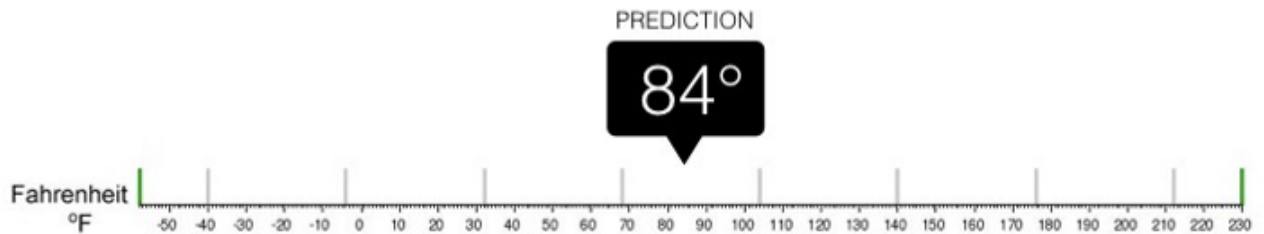
simply|learn

©SimplyLearn. All rights reserved.



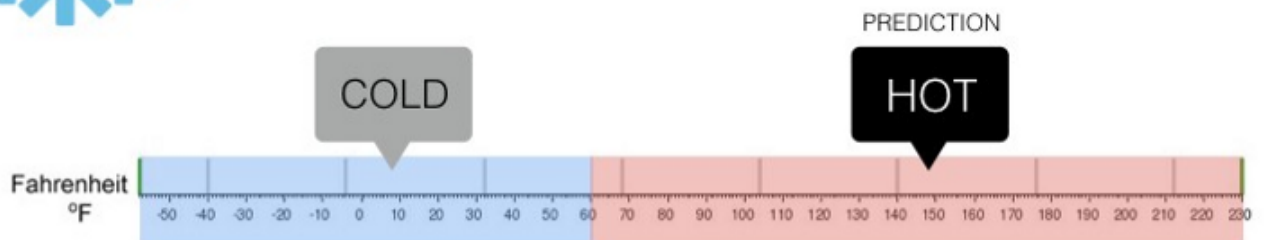
Regression

What is the temperature going to be tomorrow?



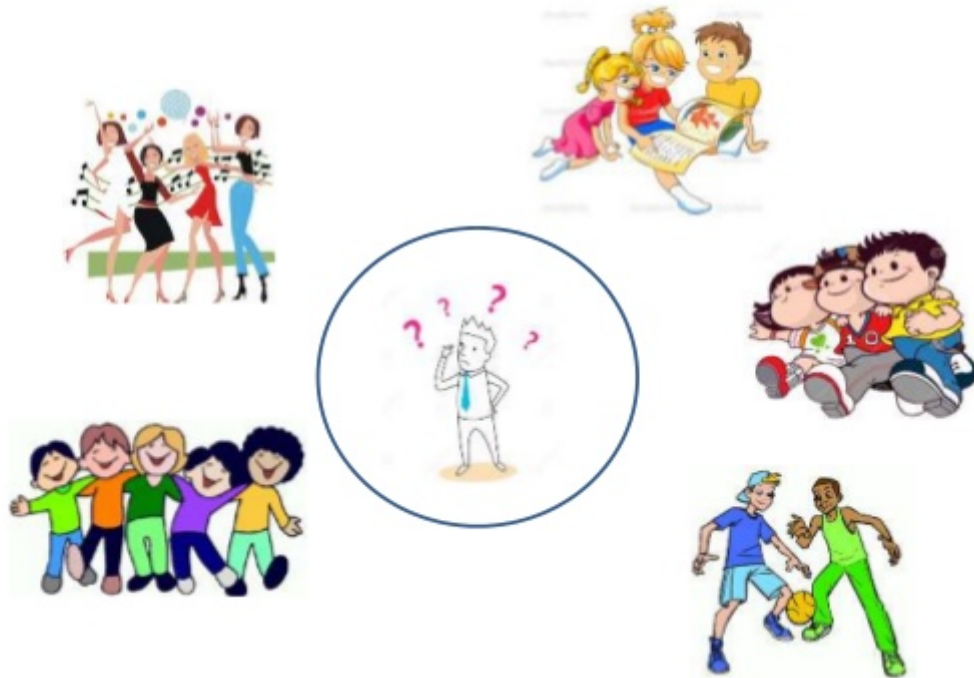
Classification

Will it be Cold or Hot tomorrow?



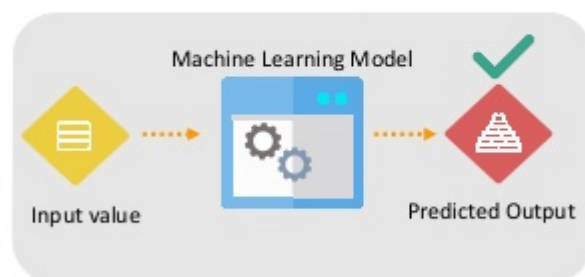
Simple Analogy..

- Tell me about your friends(*who your neighbors are*) and *I will tell you who you are*.



3

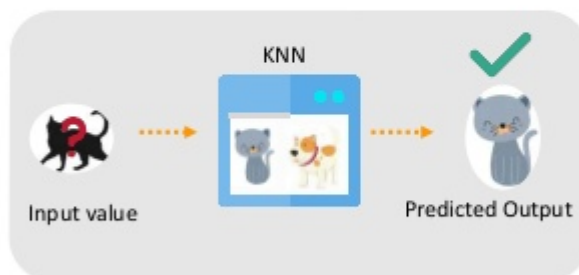
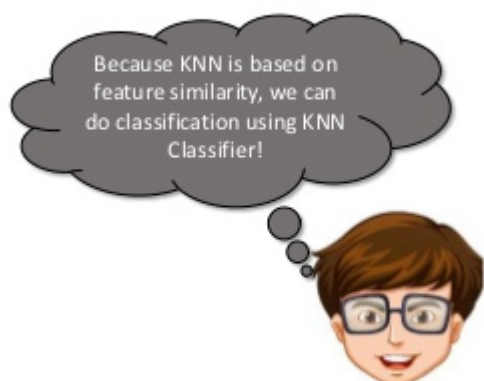
Why KNN?



©Simplilearn. All rights reserved.

simplilearn

Why KNN?

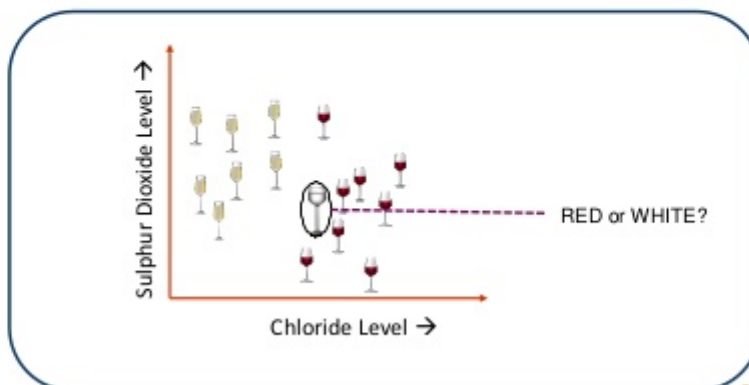


©Simplilearn. All rights reserved.

simplilearn

What is KNN Algorithm?

KNN stores all available cases and classifies new cases based on a similarity measure

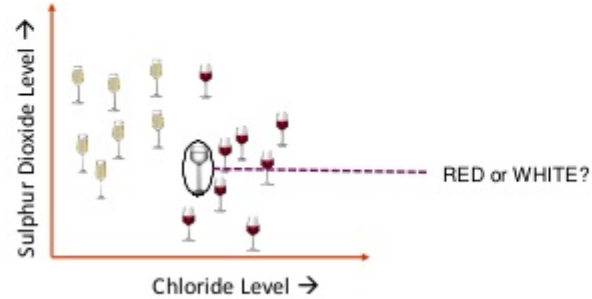


©Simplilearn. All rights reserved.

simplilearn

What is KNN Algorithm?

But, what is K?

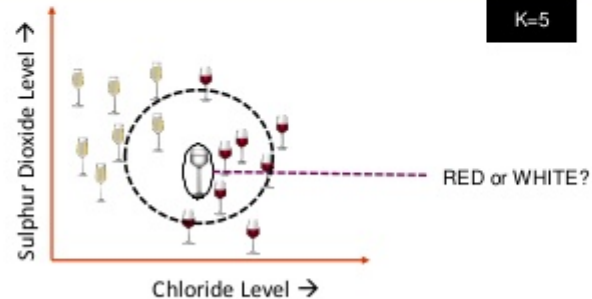


©Simplilearn. All rights reserved.

simplilearn

What is KNN Algorithm?

A data point is classified by majority votes from its 5 nearest neighbors

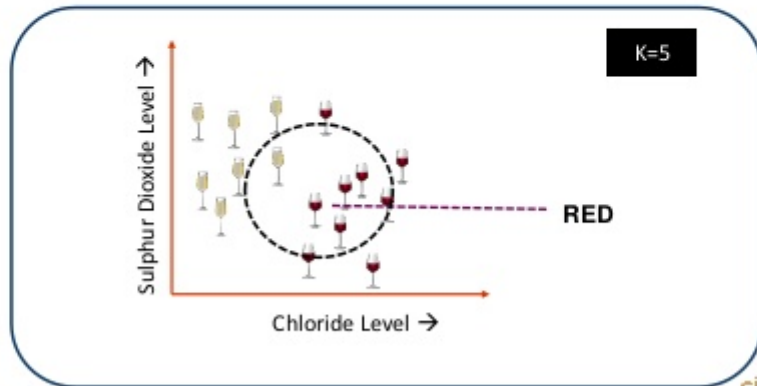


©Simplilearn. All rights reserved.

simplilearn

What is KNN Algorithm?

Here, the unknown point would be classified as red, since 4 out of 5 neighbors are red



©Simplilearn. All rights reserved.

simplilearn

How do we choose 'k'?



©Simplilearn. All rights reserved.

simplilearn

How do we choose the factor 'k'?

KNN Algorithm is based on **feature similarity**: Choosing the right value of k is a process called parameter tuning, and is important for better accuracy

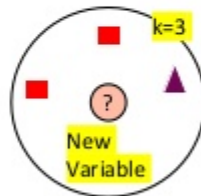


©Simplilearn. All rights reserved.

simplilearn

How do we choose the factor 'k'?

KNN Algorithm is based on **feature similarity**: Choosing the right value of k is a process called parameter tuning, and is important for better accuracy



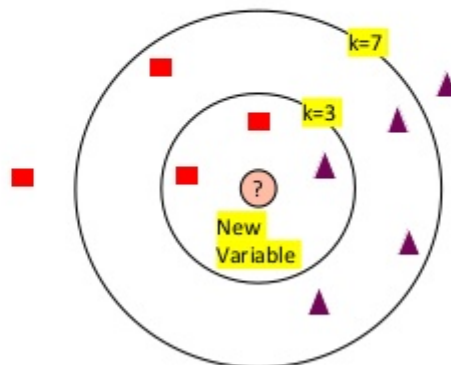
So at $k=3$, we can classify '?' as ■

©Simplilearn. All rights reserved.

simplilearn

How do we choose the factor 'k'?

KNN Algorithm is based on **feature similarity**: Choosing the right value of k is a process called parameter tuning, and is important for better accuracy



But at $k=7$, we classify '?' as



©Simplilearn. All rights reserved.

simplilearn

How do we choose the factor 'k'?

KNN Algorithm is based on feature similarity: Choosing the right value of k is a process called parameter tuning, and is important for better accuracy



So at $k=3$, we can classify '?' as



©Simplilearn. All rights reserved.

simplilearn

How do we choose the factor 'k'?

To choose a value of k:

\sqrt{n} , where n is the total number of data points

Odd value of K is selected to avoid confusion between two classes of data

©Simplilearn. All rights reserved.

simplilearn

How do we choose the factor 'k'?

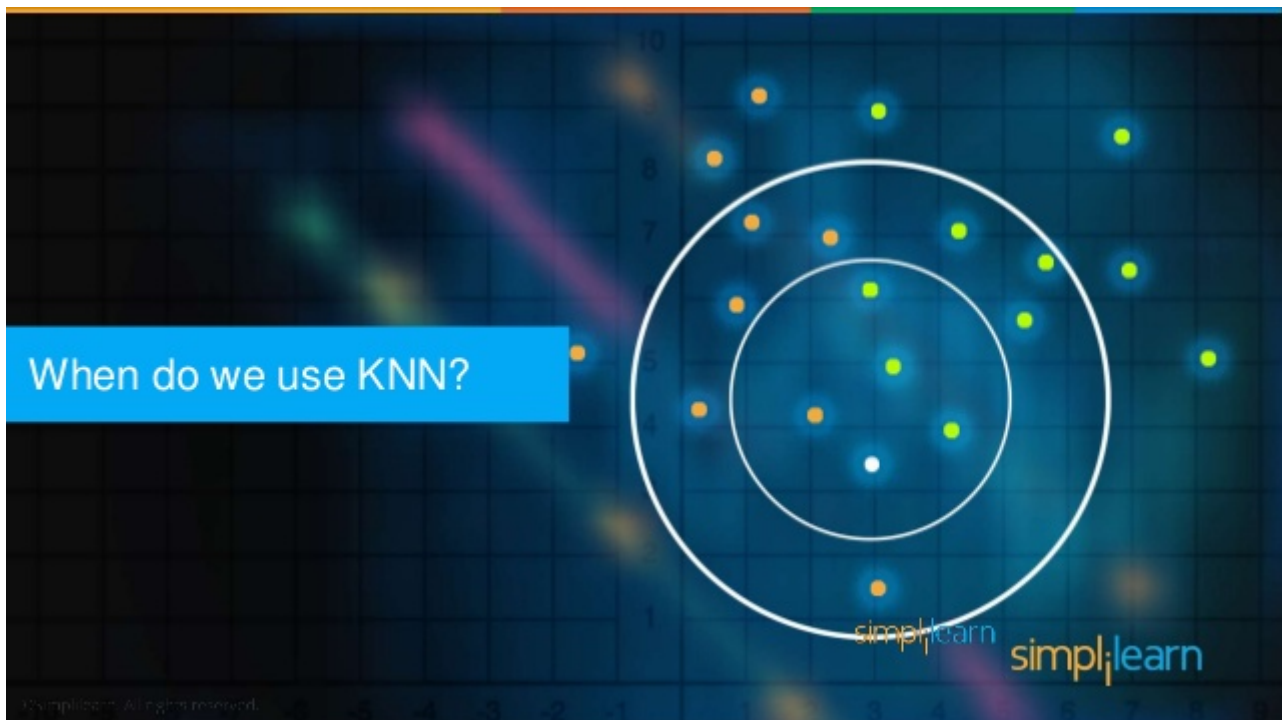
Higher value of k has lesser chance of error

\sqrt{n} , where n is the total number of data points

Odd value of K is selected to avoid confusion between two classes of data

©Simplilearn. All rights reserved.

simplilearn



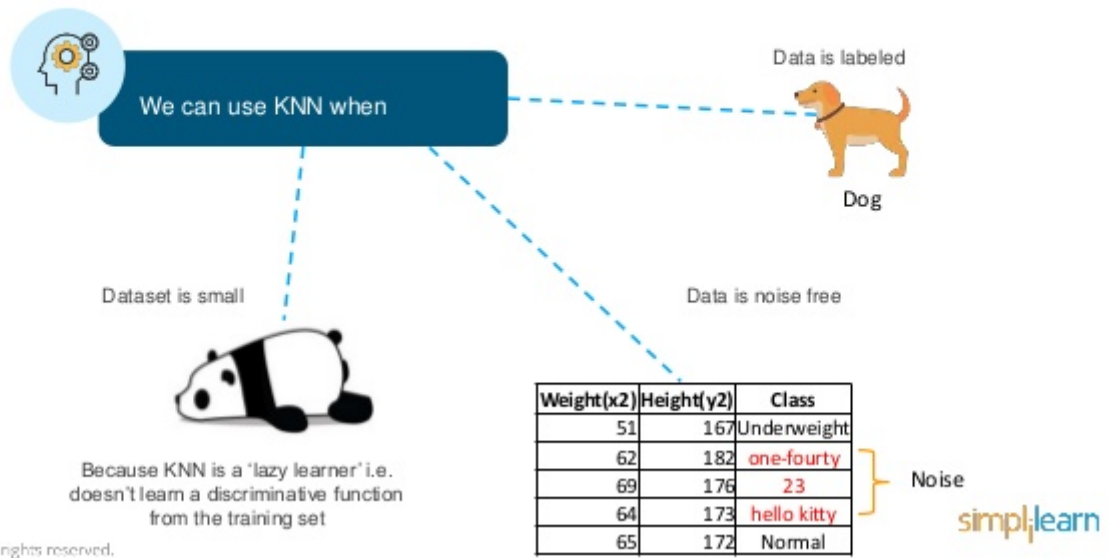
When do we use KNN Algorithm?



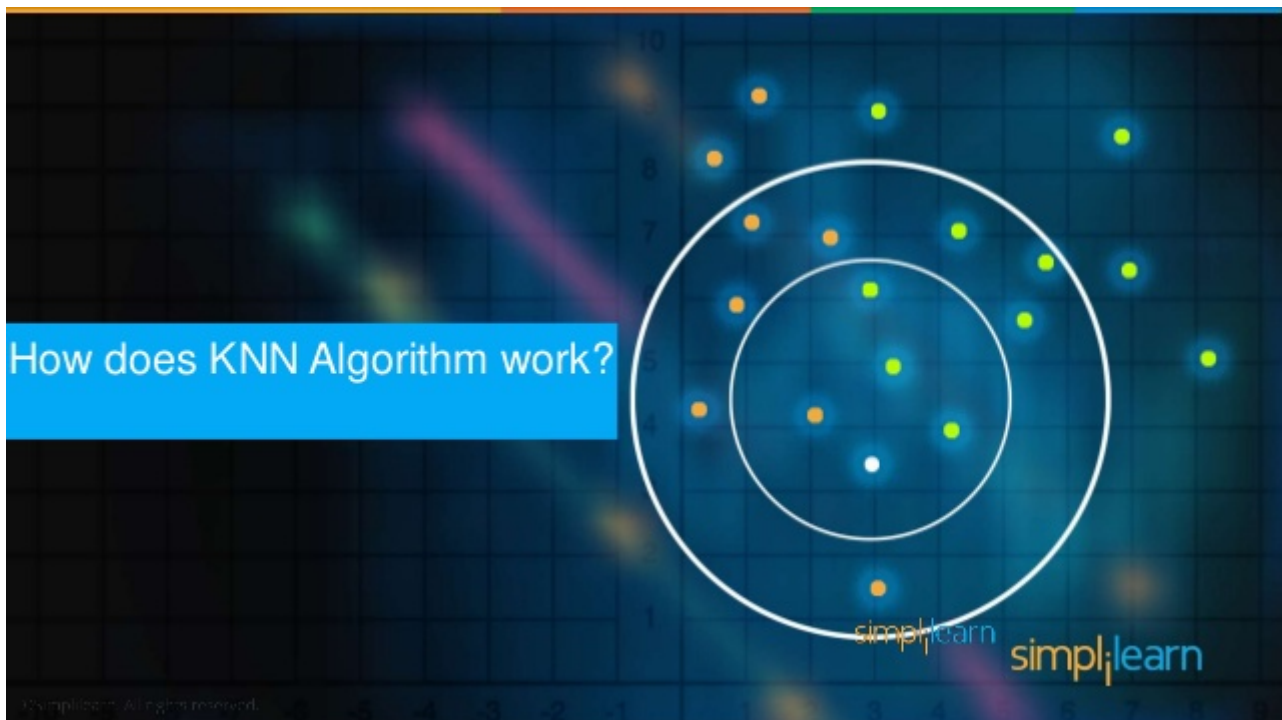
©Simplilearn. All rights reserved.

simplilearn

When do we use KNN Algorithm?



How does KNN Algorithm work?



How does KNN Algorithm work?



Consider a dataset having two variables: height (cm) & weight (kg) and each point is classified as Normal or Underweight

Weight(x2)	Height(y2)	Class
51	167	Underweight
62	182	Normal
69	176	Normal
64	173	Normal
65	172	Normal
56	174	Underweight
58	169	Normal
57	173	Normal
55	170	Normal

©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?



On the basis of the given data we have to classify the below set as Normal or Underweight using KNN

57 kg	170 cm	?
-------	--------	---



Assuming, we don't know how to calculate BMI!

©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?

To find the nearest neighbors, we will calculate Euclidean distance



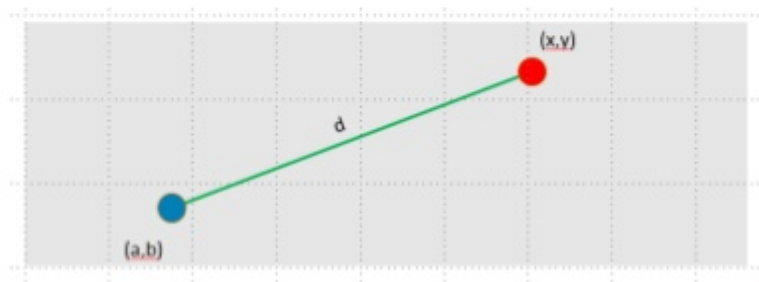
©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?

According to the **Euclidean distance** formula, the **distance** between two points in the plane with coordinates (x, y) and (a, b) is given by:

$$\text{dist}(d) = \sqrt{(x - a)^2 + (y - b)^2}$$

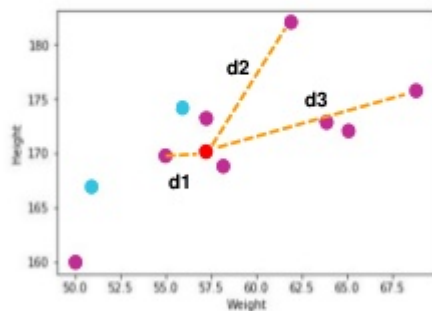


©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?

Let's calculate it to understand clearly:



$$\text{dist}(\mathbf{d1}) = \sqrt{(170-167)^2 + (57-51)^2} \approx 6.7$$

$$\text{dist}(\mathbf{d2}) = \sqrt{(170-182)^2 + (57-62)^2} \approx 13$$

$$\text{dist}(\mathbf{d3}) = \sqrt{(170-176)^2 + (57-69)^2} \approx 13.4$$

Similarly, we will calculate Euclidean distance of unknown data point from all the points in the dataset

● Unknown data point

©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?

Hence, we have calculated the Euclidean distance of unknown data point from all the points as shown:

Where $(x1, y1) = (57, 170)$ whose class we have to classify

Weight(x2)	Height(y2)	Class	Euclidean Distance
51	167	Underweight	6.7
62	182	Normal	13
69	176	Normal	13.4
64	173	Normal	7.6
65	172	Normal	8.2
56	174	Underweight	4.1
58	169	Normal	1.4
57	173	Normal	3
55	170	Normal	2

©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?

Now, lets calculate the nearest neighbor at $k=3$

Weight(x2)	Height(y2)	Class	Eudidean Distance
51	167	Underweight	6.7
62	182	Normal	13
69	176	Normal	13.4
64	173	Normal	7.6
65	172	Normal	8.2
56	174	Underweight	4.1
58	169	Normal	1.4
57	173	Normal	3
55	170	Normal	2

$k = 3$



57 kg	170 cm	?
-------	--------	---

©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?

Now, lets calculate the nearest neighbor at $k=3$

Weight(x2)	Height(y2)	Class	Eudidean Distance
51	167	Underweight	6.7
62	182	Normal	13
69	176	Normal	13.4
64	173	Normal	7.6
65	172	Normal	8.2
56	174	Underweight	4.1
58	169	Normal	1.4
57	173	Normal	3
55	170	Normal	2

We have $n=10$,
And $\text{sqrt}(10)=3.1$
Hence, we have taken $k=3$



57 kg	170 cm	?
-------	--------	---

©Simplilearn. All rights reserved.

simplilearn

How does KNN Algorithm work?



Class	Euclidean Distance
Underweight	6.7
Normal	13
Normal	13.4
Normal	7.6
Normal	8.2
Underweight	4.1
Normal	1.4
Normal	3
Normal	2



So, majority neighbors are pointing towards 'Normal'

Hence, as per KNN algorithm the class of (57, 170) should be 'Normal'

©Simplilearn. All rights reserved.

simplilearn

Recap of KNN



Recap of KNN

- A positive integer k is specified, along with a new sample
- We select the k entries in our database which are closest to the new sample
- We find the most common classification of these entries
- This is the classification we give to the new sample

©Simplilearn. All rights reserved.

simplilearn