1

00:00:00,610 --> 00:00:04,690

[Auto-generated transcript. Edits may have been applied for clarity.]

The next algorithm is called k nearest neighbors or k n n.

2

00:00:04,810 --> 00:00:09,430

Although it's considered a machine learning algorithm. It does not have a training phase.

3

00:00:09,760 --> 00:00:15,970

It's known as a nonparametric algorithm, meaning it doesn't involve learning model parameters during training.

4

00:00:16,330 --> 00:00:20,020

Instead, it operates directly on the data during prediction.

5

00:00:20,450 --> 00:00:24,010

K n is used for both classification and regression.

6

00:00:24,340 --> 00:00:32,050

It's considered a machine learning algorithm because it's widely applied in practice, even though it doesn't involve a traditional learning process.

7

00:00:32,530 --> 00:00:36,579

The idea is simple. Given a new data point, for example,

8

00:00:36,580 --> 00:00:44,530

in a classification task K and finds the k closest data points in the training set and predicts the class based on them.

9

00:00:45,610 --> 00:00:51,160

Consider a case where you want to classify a person as male or female based on their height.

10

00:00:51,520 --> 00:00:54,820

One approach would be to use logistic regression.

11

00:00:55,270 --> 00:00:57,159

Another approach is to use.

12

00:00:57,160 --> 00:01:04,480

Can you find the five people with the closest heights to the new person, and assign the class that is most frequent among those five?

13

00:01:05,590 --> 00:01:10,750

For instance, suppose the new person has a height of 179cm.

14

00:01:11,260 --> 00:01:26,440

You look at your data set and find the five closest heights 181cm, 180.5cm, 179cm, 178.5cm, and 177.8cm.

15

00:01:26,710 --> 00:01:34,390

Then check the gender of those five individuals. If the majority are male, you assign the label male to the new person.

16

00:01:34,690 --> 00:01:38,770

The parameter k in this case five is user defined.

17

00:01:39,400 --> 00:01:45,130

Tan is also used for regression. For example, suppose you want to predict chest circumference based on height.

18

00:01:45,490 --> 00:01:48,490

A new person comes in with a height of 180cm.

19

00:01:48,790 --> 00:01:56,560

You find the three people in your data set with the closest heights, whose chest circumferences are 52, 55, and 51cm.

20

00:01:57,040 --> 00:02:06,160

You average these values 52 plus 55, plus 51 divided by three, which equals to 52.7cm.

21

00:02:06,250 --> 00:02:11,200

So the predicted chest circumference for the new person is 52.7cm.

22

00:02:11,830 --> 00:02:20,410

This is the essence of K-nearest neighbors. It's simple yet effective and remains widely used for both classification and regression tasks.