1

00:00:00,570 --> 00:00:06,210

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

Decision trees are a machine learning model that uses a series of yes no questions to partition data.

2

00:00:06,240 --> 00:00:12,690

For example, you can use a decision tree to classify whether a person is low risk or high risk for a heart attack.

3

00:00:13,020 --> 00:00:19,440

You start with a root node that asks a yes no question, such as is the age less than 18?

4

00:00:19,920 --> 00:00:24,210

If yes, you check the weight. If it's less than 60, it's low risk.

5

00:00:24,450 --> 00:00:31,950

If more than 60, it's high risk. If the age is between 18 and 30, it is classified as low risk and so on.

6

00:00:32,190 --> 00:00:40,950

When a new data point comes in with attributes like age, smoker status, and weight, the model uses the learned structure to classify the individual.

7

00:00:41,460 --> 00:00:49,290

For instance, a 31 year old smoker weighing 60kg will be classified by traversing the tree based on these attributes.

8

00:00:49,860 --> 00:00:59,280

During training, the algorithm learns the splits or the parameters of these questions using a loss function based on maximizing information gain.

9

00:00:59,880 --> 00:01:06,750

For example, if you have a data set with attributes such as outlook, windy, temperature, humidity,

10

00:01:06,990 --> 00:01:14,130

and a target indicating if you can play outside, the algorithm evaluates which attribute to split on first.

11

00:01:14,370 --> 00:01:19,410

It tests whether to split on outlook, windy, humidity, or temperature.

12

00:01:20,160 --> 00:01:24,720

If splitting on outlook results in an information gain of 0.247,

13

00:01:24,930 --> 00:01:31,080

and this is higher than splitting on other attributes, the algorithm will choose outlook as the first split.

14

00:01:31,170 --> 00:01:34,830

Information gain measures how well a split creates pure nodes.

15

00:01:35,130 --> 00:01:40,090

For instance, splitting on outlook and getting overcast results in all.

16

00:01:40,110 --> 00:01:45,150

Yes, for playing outside it creates a pure node indicating a good split.

17

00:01:45,540 --> 00:01:52,230

The tree continues branching on the next attributes, selecting splits that maximize purity and information gain.

18

00:01:52,530 --> 00:01:59,940

The algorithm recursively tests different splits, learning the parameters that lead to the best classification performance.

19

00:02:00,510 --> 00:02:05,400

The goal is to create a tree where each split results in pure or nearly pure nodes.

20

00:02:05,580 --> 00:02:14,490

Improving classification accuracy. Ensemble algorithms enhance decision trees by combining multiple trees, resulting in a powerful model.

21

00:02:14,790 --> 00:02:21,510

One such ensemble method is random forests, where multiple decision trees vote on the classification of new data.

22

00:02:22,020 --> 00:02:27,210

In random forests, each tree is trained on the data set using the information gain method,

23

00:02:27,360 --> 00:02:34,170

but with random exclusions of certain nodes or attributes, resulting in different trees with different structures.

24

00:02:34,590 --> 00:02:40,200

Each tree votes on the class, and the majority vote determines the final classification.

25

00:02:40,500 --> 00:02:44,370

Random forests can be used for both classification and regression.

26

00:02:44,790 --> 00:02:49,020

They are effective because the different trees capture different aspects of the data,

27

00:02:49,290 --> 00:02:53,820

and their combined voting mechanism leads to more robust and accurate predictions.