1

00:00:00,560 --> 00:00:05,300

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

Let's talk about the foundational machine learning algorithm, linear regression.

2

00:00:05,630 --> 00:00:09,950

It's widely studied in Stem and mathematically oriented courses.

3

00:00:10,340 --> 00:00:20,990

Despite its simplicity, linear regression is a true machine learning algorithm and clearly illustrates what machine learning is in its simplest form.

4

00:00:21,200 --> 00:00:25,220

It determines a linear relationship between two variables.

5

00:00:25,550 --> 00:00:31,070

For example, the input could be years of experience and the goal is to predict salary.

6

00:00:31,550 --> 00:00:34,970

Essentially, you're fitting a line to the data.

7

00:00:36,790 --> 00:00:45,549

While there are other types of regression. Linear regression specifically fits a line to the data using a learning algorithm that

8

00:00:45,550 --> 00:00:50,890

minimizes the sum of squared distances between data points and the regression line.

9

00:00:52,740 --> 00:01:03,000

Suppose you start with a randomly drawn line. The goal is to adjust the slope beta and the intercept epsilon so the line fits the data better.

10

00:01:05,460 --> 00:01:10,650

The initial line may not fit well, so beta and epsilon are tweaked iteratively.

11

00:01:11,310 --> 00:01:17,400

You begin with random values and update them to reduce the discrepancy between the predicted line and actual data points.

12

00:01:18,090 --> 00:01:23,190

This discrepancy is measured as the squared difference between each point and the line.

13

00:01:23,730 --> 00:01:28,200

Every training algorithm in machine learning is an optimization process.

14

00:01:28,650 --> 00:01:35,130

You optimize parameters beta and epsilon so that these squared differences are minimized.

15

00:01:35,580 --> 00:01:40,620

We square the differences because we don't care whether the prediction error is positive or negative.

16

00:01:40,860 --> 00:01:45,750

Squaring ensures all errors contribute positively. Why minimize these errors?

17

00:01:46,200 --> 00:01:49,740

Because we want the model to generalize well to unseen data.

18

00:01:50,190 --> 00:01:55,110

Suppose you're hiring someone and want to estimate their salary based on ten years of experience.

19

00:01:55,530 --> 00:02:02,280

Even if that exact data point wasn't in your training set, you can still make a prediction using the fitted line when squared.

20

00:02:02,280 --> 00:02:09,960

Differences are small. The fit is better. This process is formalized using a loss function for linear regression.

21

00:02:10,140 --> 00:02:14,070

The loss function is mean squared error or MSE for short.

22

00:02:14,490 --> 00:02:19,320

During training, the algorithm minimizes this function to find the best fitting line.

23

00:02:19,800 --> 00:02:23,550

It's important to distinguish between parameters and hyperparameters.

24

00:02:23,880 --> 00:02:29,280

Parameters are values learned during training. Beta slope and epsilon intercept.

25

00:02:29,680 --> 00:02:33,990

Hyperparameters are configuration settings that control the learning process.

26

00:02:34,380 --> 00:02:39,450

For example, the learning rate determines how quickly the model updates its parameters.

27

00:02:39,900 --> 00:02:44,370

Hyperparameters are set before training and require experimentation to tune.

28

00:02:44,880 --> 00:02:52,050

This example uses a one dimensional model where one feature experience predicts one output salary.

29

00:02:52,560 --> 00:02:55,920

Dimensionality refers to the number of input features.

30

00:02:56,340 --> 00:03:01,890

If your dataset includes only one feature and one target, it's one dimensional regression.

31

00:03:02,400 --> 00:03:05,910

You can increase dimensionality by adding more variables.

32

00:03:06,180 --> 00:03:12,330

For example, to predict height, you might use arms length, biological sex, and athletic status.

33

00:03:12,480 --> 00:03:20,070

These variables form a multidimensional model or three dimensional in this case that better captures variation in height.

34

00:03:20,610 --> 00:03:27,960

You'll still have an intercept. And in higher dimensions, the fitted model is a linear combination of features.

35

00:03:28,680 --> 00:03:39,660

While we can graph up to three dimensions in n dimensional problems, the model is still mathematically aligned, or more precisely, a hyperplane.

36

00:03:40,290 --> 00:03:49,920

Linear regression introduces core concepts like parameters, hyperparameters, dimensionality, training algorithms, and loss functions.

37

00:03:50,610 --> 00:03:54,420

To summarize, the loss function is what you minimize during training.

38

00:03:54,840 --> 00:03:58,710

For linear regression, the default is mean squared error.

39

00:03:58,980 --> 00:04:04,140

You sum the squared differences between predicted and actual values for all data points.

40

00:04:04,830 --> 00:04:13,650

The smaller the sum, the better the fit. Although mean squared error is common, other loss functions may be used depending on the task.