SUPERVISED MACHINE LEARNING

1. Supervised machine learning

00:00 - 00:00

In this video, we'll dive into machine learning!

2. Data science workflow

00:00 - 00:00

As we learned previously, machine learning is a set of methods for making predictions based on existing data, hence it belongs in the last step of the workflow.

3. What is supervised machine learning?

00:00 - 00:00

Supervised machine learning is a subset of machine learning where the existing data has a specific structure: it has labels and features. More on that later. Examples of its abilities includes recommendation systems, diagnosing biomedical images, recognizing hand-written digits, and predicting customer churn. Let's define these new terms with a case study.

4. Case study: churn prediction

00:00 - 00:00

Suppose we have a subscription business and want to predict whether a given customer is likely to stay subscribed or cancel their subscription, also known as churn.

5. Case study: churn prediction

00:00 - 00:00

First, we need some training data to build our model off of. This would be historical customer data.

6. Case study: churn prediction

00:00 - 00:00

Some of those customers will have maintained their subscription, while others will have churned. We eventually want to be able to predict the label for each customer: churned or subscribed.

7. Case study: churn prediction

00:00 - 00:00

We'll need features to make this prediction. Features are different pieces of information about each customer that might affect our label. For example, perhaps age, gender, the date of last purchase, or household income will predict cancellations. The magic of machine learning is that

we can analyze many features all at once. We use these labels and features to train a model to make predictions on new data.

8. Case study: churn prediction

00:00 - 00:00

Suppose we have a customer who may or may not churn soon. We can collect feature data on this customer, such as age, or date of last purchase.

9. Case study: churn prediction

00:00 - 00:00

We can feed this data into our trained model

10. Case study: churn prediction

00:00 - 00:00

and then, our trained model will give us a prediction. If the customer is not in danger of churning, we can count on their revenue for another month! If they are in danger of churning, we can reach out to them to try to keep them subscribed.

11. Supervised machine learning recap

00:00 - 00:00

Let's recap. Machine learning makes a prediction based on data. In supervised machine learning, that data has two characteristics: features and labels. Labels are what we want to predict, like the customer churning. Features are data that might help predict the label, such as profession or date of last purchase. Once we have the features and labels, we train a model and use it to make predictions on new data.

12. Model evaluation

00:00 - 00:00

After training a model, how do we know if it's any good? It's always good practice not to allocate all of your historical data for training. Withheld data is called a test set and can be used to evaluate the goodness of the model. In our example, we could ask the model to predict whether a set of customers would churn, and then measure the accuracy of our prediction.

13. Model Evaluation

00:00 - 00:00

For example, let's say we're testing our model on our test set made up of 1000 customers, where only 30 of the customers have actually churned.

14. Model Evaluation

00:00 - 00:00

We put that test data into our newly trained model and it predicts that all the customers remain.

15. Model Evaluation

00:00 - 00:00

If we calculate the overall accuracy of that model, it technically has a high accuracy of 97% because it was correct on 970 of the 1000 customers. This is despite never correctly labeling a customer churning. Checking both outcomes is important for rare events. Only by examining the accuracy of each label do we get 0% accuracy at predicting churn when churn was the actual outcome. This model is not useful to use in its current state, so we'll have to re-train it with different parameters or more data.

16. Let's practice!

00:00 - 00:00

You learned how supervised learning uses features and labels and how to be cautious with model accuracy. Time for practice!