Twitter Thread by Santiago



Santiago
@svpino

11 key concepts of Machine Learning.

— Supervised Learning Edition —

■■

■

Before starting, remember that, if you follow me, one of your enemies will be immediately destroyed (and you'll get to read more of these threads, of course.)

And if you don't follow me, well, you just hurt my feelings.

■

1. Labels

(Also referred to as "y")

The label is the piece of information that we are predicting.

For example:

- the animal that's shown in a picture
- the price of a house
- whether a message is spam or not

■

2. Features

(Also referred to as "x")

These are the input variables to our problem. We use these features to predict the "label."

For example:

- pixels of a picture
- number of bedrooms of a house
- square footage of a house

■

3. Samples

(This is also known as "examples.")

A sample is a particular instance of data (features or "x.") It could be "labeled" or "unlabeled."

■

4. Labeled sample

Labeled samples are used to train and validate the model. These are usually represented as (x, y), where "x" is a vector containing all the features, and "y" is the corresponding label.

For example, a labeled sample could be:

([3, 2, 1500], 350000)

5. Unlabeled sample

Unlabeled samples contain features, but they don't contain the label: (x, ?)

We usually use a model to predict the labels of unlabeled samples.

■

6. Model

A model defines the relationship between features and the label.

You can think of a model as a set of rules that, given certain features, determines the corresponding label.

For example, given the # of bedrooms, bathrooms, and square footage, we get the price.

■

7. Training

Training is a process that builds a model.

We show the model labeled samples during training and allow the model to gradually learn the relationships between features and the label.

■

8. Validation

Validation is the process that lets us know whether a model is any good.

Usually, we run a set of (unseen) labeled samples through a model to ensure that it can predict the labels.

■

9. Inference

Inference is the process of applying a trained model to unlabeled samples to obtain the corresponding labels.

In other words, "inference" is the process of making predictions using a model.

■

10. Regression

A regression model predicts continuous values, for example:

- the value of a house
- the price of a stock
- tomorrow's temperature

■

11. Classification

A classification model predicts discrete values, for example:

- the picture is showing a dog or a cat
- the message is spam or not
- the forecast is sunny or overcast