Section 1.1: Introduction to Machine Learning

Monday, September 5, 2022 17:14

Consider the price of a car listed on a car classifieds website; imagine there is a person who wants to sell their car.

This person would photograph the car and upload it to the website in order to set their asking price for the vehicle. This is difficult, because you don’t want to put a price that is too high, because nobody would want to buy the car; conversely you don't want to put a price that is too low because you could be leaving money on the table (in other words, you could have potentially sold it for a higher price than you set).

So, you want to find a price that is optimal.

Machine Learning (ML) can help the user find the best price to list the car at based on outlined and specified parameters.

Returning to the car example, we can assess the car's price, its age, its manufacturer, the mileage, and other bits of information about the car; take the information to a dealership and an expert looks at the car. Then the expert considers the provided information, in order to determine where exactly to price the car.

Experts learn how to make such determinations from data. The data can be internal or external to the dealership; for example, they can look at data from other dealerships. Then they inspect this data for patterns, and they know that certain correlations exist based on the characteristics. Some of these correlations could be simple things such as an older car being less expensive than a newer car, or a car with more mileage is less expensive than a car with less mileage, among other things.

The point of this example is to say that if an expert can make these determinations, then so can a model. If we have a dataset with all these characteristics of a car and their prices, we can take all of this data, and put all of the information there and the model will learn the pattern itself.

The essence of machine learning is that you take data and the model extracts a pattern from the data. This way, we can also replicate what is to be learned from the data by experts.

Back to the car example, we have, "features" which is what we know about the cars(year, make, mileage, etc.); then we have the "target" which is what we want to predict (in this case, the price).

Then we execute: "features" + "target" = model. Then we use the model to make price predictions for the cars about which we are uncertain to price. We input the data pertaining to the features we know about the car into the model, so the model could output the prediction price.

The model is not able to correctly predict the price of every car, but the predictions are usually correct number average. So a car with specific features [x, y, z] would cost a certain amount, A; versus another car with specific features [t, u, v] would cost a different amount B; where B may or may not equal to A.

In other words, the output varies based on the specified data.

So, once we have the model, we can use it to help our user to determine the price of the car. We take the information we have about the features, or data, and input the data into the model, and

whatever the model produces is the price. Now, the user does not have to concern themselves with setting the price, the model will produce a price and the user can make their own discretionary adjustments.

Machine learning is the process of extracting patterns from data; and data is usually of two types (features and targets).

The concept of ML is depicted with an example of predicting the price of a car. The ML model learns from data, represented as some **features** such as year, mileage, among others, and the **target** variable, in this case, the car's price.

Notes by Community:

Then, the model extracts patterns from the data and predicts the price of cars that were not considered in the training data.

In summary, ML is a process of **extracting patterns from data**, which is of two types:

- features (information about the object) and
- target (correct predictions).

Therefore, new features are presented to the model, and it makes **predictions** considering learned information.