Module 2 Transcript

Video 2.2: What is machine learning?

Let's start with a fundamental question, what is machine learning actually about? It turns out that a large part of machine learning is about a seemingly very simple problem. It is about understanding the relationship between a number of input variables, which are denoted here by 'X_1' up to 'X_P' and an output variable which is denoted by Y here. Now learning this relationship between these input variables and the output variable turns out to be rather tricky for two reasons. The first one is we only have limited data points that allow us to learn this relationship between the input variables and output variables from.

So, if we describe the relationship between the input variables and the output variables as a function 'f' that we would like to learn, there is limited data available to learn 'f' from. The second complication is the presence of noise. Even if we had as many data points as we wanted to, we will never be able to learn the function 'f' with absolute certainty because the connection between the input variables 'X_1' to 'X_P' and Y is the stochastic one. It is affected by noise, which could be other input variables that we have left out in our model, that we haven't measured for example.

Or it could be that the underlying true relationship between the input variables and the output variable is an inherently stochastic one. By the way, the input variables come in many different names in machine learning. We sometimes call them also independent variables or predictors, features or fields. Likewise, for the output variable, we sometimes use the term dependent variable or response variable, target variable, or outcome variable. Now let's explore why we may want to learn the relationship between input variables and the output variable in practice.
It turns out that there are two major reasons for this: one is to do forecasting; the other one is to do inference. Let's look at these two reasons in turn. In forecasting, we want to be able to predict the value of the outcome variable 'Y' from the input variables from the values of the input variables 'X_1' to 'X_P' on data that we have not yet seen.

On data where we have only seen the input variables, but not the output variable. Let's have a look at a couple of examples. Imagine for example you want to verify or determine whether a patient has cancer or not. In that case, you may get a blood sample. The blood sample has many different characteristics which could be captured by input variables 'X_1' to 'X_P'. The output variable would be a binary variable which tells us whether the patient has a certain type of cancer or not.
Or it could be what we call later on a categorical variable which tells us which type of cancer the patient has, or hopefully none at all.

A different example would be to detect fraudulent transactions from let's say tax claims or expense claims. In that case, the input variables would be certain properties of the let's say the expense claim, the output variable would be a binary variable which says this expense claim is fraudulent or not.

There are many other applications like that. The key to forecasting is that the sole the sole motivation for us is to predict with high accuracy what the value of the output variable is for given variables of the input variables. The function 'f' that we estimate could be arbitrarily complex here. In practice, for example, we use neural networks in many cases. These are predictors, we will see them later on. These are predictors that work often extremely well, but they lead to extremely complicated functional relationships 'f'. In forecasting we don't mind; all we care about is prediction accuracy.

The other class of problems that often arise in machine learning are so called inference problems. In inference problems, we actually do care about our explanation for the functional relationship 'f'. We don't want our estimation for this functional relationship 'f' which we often call f-hand. We don't want that explanation to be too complicated because we want to actually understand what the relationship is between input variables and output variables. Let's have a look at some examples. Imagine, for example, we want to predict the sales of different types of marketing campaigns depending on the medium that we use.

For example, it could be a TV campaign or newspaper campaign or radio campaign. In that case, we're not just interested in predicting what the increase in sales is. We want to understand how the increase in sales can be attributed to different campaign types because we want to then take a decision: what is our optimal marketing mix? Another example would be the prediction of house prices. In many cases, we do not just want to predict the price of a house based on properties of a house, such as is it a Victorian bill? Does it have a river view or things like that? But we also want to understand, what is the contribution of the river view to the house price? Those are inference problems where we do not just care about prediction accuracy, but we also care about a simple model that can be interpreted by a human decision-maker.