Introduction to Machine Learning

Slide 1

In this video, we will define machine learning and discuss the various branches of the discipline.

Machine learning is a subset of artificial intelligence which is defined as the ability of a computer or computer-controlled machine to perform tasks that require human intelligence.

Some examples include having a conversation, writing a book, driving a car, and proving a mathematical theorem.

Machine learning is branch of AI that gives computers the capability to learn without being explicitly programmed.

Slide 2

What exactly is an explicit program? Before machine learning, computers had to be given precise instructions to achieve any task.

Let’s look at a simple example. Assume our goal is to write an algorithm that can tell us if a number is even or not.

How could we achieve this? We would write code to perform the following: given a number, divide it by two, and if the remainder is 0, return “Yes”, otherwise return “No”.

Here is an example program execution. We are given the number 6, our program executes the logic we just discussed, and we get a “Yes”.

What are the benefits of programming in this way? One is that we will get the correct output on every execution. The downside of this method is that all rules to accomplish a task must be known in advance.

For this simple problem, that isn’t an issue. But it can be a challenge for complicated tasks.

Slide 3

When writing explicit programs, the typical project workflow starts but studying the problem and determining the rules needed to achieve the desired results.

Then we begin an iterative process of applying our rules to a given data set, analyzing the errors, and updating our rules until we achieve decent results.

Once this is done, we launch our algorithm into production where it is applied to new data.

Then we monitor the results and update our rules as needed.

This process can be very labor intensive for complex tasks that require thousands or even millions of rules.

Slide 4

Machine learning takes a different approach. These algorithms use vast amounts of data to discover patterns and relationships without relying on a set of predetermined rules.

If we go back to our simple example of determining whether a number is even or odd, we would pass a labeled data set to the machine learning algorithm and obtain a prediction function.

Each row of our data set represents an observation with a column that has the correct answer and another one with the input data value. Machine learning algorithms will use the patterns in this data to find a function that can generate predictions on new data values.

If we use this prediction function on new data, many of them will generate a correct prediction. However, things will usually not be 100 percent accurate.

The benefit of this approach is that we do not have to know all the rules in advance. But the challenge will be prediction error. The goal of machine learning is to find a function that will generate the smallest amount of errors on new data values.

Slide 5

A machine learning process differs from explicit programming. As before, we start by studying the problem we need to solve.

But instead of writing rules, we obtain data and train our machine learning algorithm.

This is usually done iteratively, where predictions functions are found, their errors get analyzed, and the function get updated or tuned as practitioners describe it.

Once a suitable prediction function is found, it gets launched into production on new data sources.

As before, the errors on new data are monitored and the algorithm gets re-trained on a periodic basis.

The key takeaway from this process, is that we do not have to worry about discovering all the rules necessary to perform a task.

Slide 6

Let’s compare these two ways of programming to the task of identifying handwritten digits. This is one task that the postal service has been able to automate thanks to machine learning.

On the righthand side of the slide we have some example images of handwritten digits from the MNIST database.

How difficult is this task? For humans, it’s easy. For computers, it’s extremely difficult.

Slide 7

Imagine having to develop an explicit program to correctly identify each handwritten digit.

You would have to code the rules for every possible variation of how digits are written. For example, sometimes a 7 will have a horizontal line through the middle. Or many times the number 9 is slanted 45 degrees to the right.

This would be practically impossible. Your program would be millions of lines long.

Slide 8

To solve this problem with machine learning, we first need to encode the images into a data set that we can use to train our algorithm.

One way to do this is to partition each image into a 28 by 28 grid. This will give us 784 regions per image.

Then, within each region, we can measure the color intensity ranging from 0, which represents no color, to 255 which represents black.

If we organize this into a data set, it will look like the data on the righthand side of the slide.

Each row in this table represents a number. The correct labels for each row are stored in the Number column whereas the color intensities are stored in the Region columns.

Now we have all that we need to train machine learning algorithms to predict the correct number for each image.

Slide 9

To show machine learning in action, let's visualize how an algorithm would iteratively get trained to distinguish these digits from each other.

The video on the right is visualizing the training of a machine learning algorithm on the data set from the previous slide. Each image is color coded and plotted on a 3-dimensional graph.

The goal of this algorithm is to find an optimal way to compress the digit data from 784 columns to just 3 in such a way as to group the same digits together.

Notice that after a couple of iterations of this training process, similar digits get mapped to the same regions in this 3-dimensional space.

Once this prediction function is discovered, we can use it to predict new digit images based on where they fall in this 3-dimensional graph.

And we have been able to do this without writing a single rule! This is exactly why machine learning is so powerful!

Next, we will discuss the different types of machine learning algorithms.

Slide 10

Supervised machine learning algorithms learn prediction function from labeled training data. A common acronym for supervised learning is predictive modeling.

In this setting, we need a labeled data set for training our algorithm. One example of a labeled data set is on the righthand side of the slide and contains information on patients at a hospital.

Each row in the data represents a patient who eventually did or did not develop heart disease. The outcome is stored in the Heart Disease column. This is typically referred to as the response variable but can also be called the target or dependent variable.

Our goal might be to predict whether a new patient will develop heart disease. We can use the information in the other columns, such as age and resting blood pressure, to train a machine learning algorithm to predict the correct outcome.

These columns are typically referred to as predictor variables but also go by feature or independent variables.

The important property of supervised learning algorithms is that each row must have a known outcome.

Slide 11

There are two main branches of supervised learning. The first is regression. In regression, we are predicting a quantitative response variable.

For example, we might want to predict the selling price of a homes using features such as square footage, age, and location.

Let's say we had the following data set for 100 homes in our area. Each row in this data represents a home and the response variable is the selling price while the predictor variable is the square footage.

An example of a prediction function that was trained using regression is displayed on the right. The points in the graph represent each row of our data and the orange curve is our prediction function.

For any new homes we get, we would use this function to predict its selling price by using its square footage.

For example, we would predict that all homes with 2,500 square feet would sell for just under $150,000.

Slide 12

The other branch of supervised learning is classification. Classification methods are used to predict categorical response variables.

An example might be predicting whether a customer will purchase a product based on the time they spent on our company's homepage and the product page.

In the data set below, each row represents a customer and our response variable consists of two categories, purchased and did not purchase. We also have two predictor variables associated with each outcome.

Instead of predicting numbers, classification algorithms produce decision boundaries for all combinations of predictor values.

In the plot on the right, each point represents a customer and their associated seconds spent on the homepage and product page. Each one of these points is colored by the true outcome, whether they did or did not purchase our product.

The classification algorithm generates estimated probabilities of a purchase for each combination of predictor values. The regions in the plot are shaded by the value of these probabilities.

The black line represents all points that have an estimated probability of 0.5. This is the decision boundary.

One way to classify new customers would be to predict that they will purchase our product if their estimated probability is greater than 0.5. This corresponds to all point in the red highlighted region.

Slide 13

Unsupervised learning is another class of machine learning algorithms. In unsupervised learning, there are feature or input variables, but no labeled outcome.

This makes things more challenging because we don't have a way to tell if our predictions are correct.

In this setting, we typically want to learn the structure and relationships present among our feature variables. Common unsupervised methods include clustering and principal components analysis which we will cover towards the end of the semester.

An example from the real world might be if you have lots of data on customer purchasing behavior and you want to discover groups or segments of customers that are similar to each other. This way you could market your products to each group in slightly different ways based on their characteristics.

Another example might be if you were given a data set of length and width measurements for a group of flowers and someone asked you if there are different species present in the data.

A common unsupervised learning algorithm used to answer these type of questions is k-means clustering where rows of a data set are grouped together based on the similarity of their feature values.

A plot of the results of this algorithm are displayed on the righthand side where each cluster or group is colored based on the algorithm results.

It seems that we have different species, but there is no way to confirm this. This is the challenge I mentioned before.

This wraps up our introduction to machine learning. Next, we will begin our study of R programming!