Overview to Machine Learning

Machine learning is the product of statistics and programming coming together and allow us to make predictions and estimations based off prepared data. This field has become very important nowadays, as data becomes more abundant and accessible for everyone.

In machine learning, data is important because it gives us something to base our statistical models on. Without data, we can't validate whether our predictions are even right. Using the data, we can apply pattern recognition algorithms to create a model which can be used to predict data based on our inputs (this is where the machine learning comes in). Once we've trained the model, it's still important to be able to test our results against test data we've set aside and make sure we maintain a satisfying level of accuracy so that they may be used in real world applications. Machine learning is a subset of artificial intelligence and often uses similar techniques. Oftentimes, they both borrow from techniques from each other.

One example of modern machine learning is seen in chatbots, such as ChatGPT. These bots are trained off experiences from chatting with previous users along with data from the internet. Such a technology would not be possible today, as human language is too complex for humans to create an algorithm for. Another example is face-recognition software. The face is made up of a multitude of complex shapes. It's impossible for us to differentiate it in code on our own, but if we have algorithms which can find patterns to detect faces themselves, this becomes possible.

An observation is a sample of data, and the feature is the label for the data (basically tells what the data means). Quantitative data is numerical and has a practically infinite set of values, while qualitative data only has a finite set of values. This is important in identifying what our data needs, as certain plots and graphs are made specifically for quantitative and qualitative data. For example, a box plot is used when we have X as a qualitative and Y as a quantitative variable, while scatter plots are used when both X and Y are quantitative.

I'm personally very interested in what machine learning has to offer. Nowadays, there are several extremely impressive feats done with machine learning and artificial intelligence. With the increase of chatbots and art generators, I believe many people are looking toward AI/ML as the next big thing. This could prove very beneficial in the professional application, as I'm sure many large companies would be looking to recruit experts in this field. As for personal projects, I would be interested in creating some simulations of, let's say, self-driving cars. Teaching a car how to properly park or drive on an in-game road would be an interesting challenge.