Overview of Machine Learning

What is machine learning?

Machine learning is a technique used in computing to perform pattern recognition on sets of data. Though it is a buzzword in today's society, the idea of using deterministic means to find patterns and predict accurately is not all that new. It has been around for a long time and is used in a variety of fields. One might consider the goal of machine learning to be training a model on a given set of data. While this is not necessarily wrong, this interpretation makes room for the overfitting problem. For a given model that is shown, or trained on, a specific set of data, it can only make predictions relative to the patterns it has established about that data. Meaning, new data that the model has not *seen* could result in poorer performance even though the data it has seen gives higher performance. With this in mind, we can rephrase the goal of machine learning as follows: we would like to generalize a model such that it can make accurate predictions on data that is both seen and unseen.

Data, patterns, and accuracy, why do we care?

In order to utilize machine learning models, we need data, and often, a lot of it. Data is important to machine learning because it is the input which models feed off of in order to determine a strategy by which to make predictions. A model's ability to recognize patterns helps to form this strategy by identifying hidden correlations within the data that might not be clear from the surface. Once a strategy has been determined, the model is able to make predictions utilizing it. Depending on the accuracy of those predictions, a machine learning model could prove to be quite valuable in the real world. For example, suppose a large multinational company utilizes a machine learning model to detect patterns of fraudulent purchases ahead of time through their store. If the model is able to accurately predict events like this, the company could save a tremendous amount of money by reducing their staff as well as the likelihood of chargebacks and streamlining a significant proportion of incoming orders to the model to determine their legitimacy.

Artificial Intelligence vs. Machine Learning

Among other buzzwords today, artificial intelligence is a popular one as well. Artificial intelligence (AI) is a field involved with computer systems mimicking human intelligence. Machine learning is employed within AI. A computer makes predictions and those predictions are satisfied by a pattern recognized in data. Machine learning on the other hand is often considered as a subfield of artificial intelligence. Vast amounts of data are fit to a machine learning model which utilizes pattern recognition to generate a strategy, as to how to predict. Machine learning makes use of historical data in a variety of forms in order to create this pattern-wise strategy.

Modern machine learning

Machine learning is a technique applied to historical data so that a computer system can *learn.* A few modern examples of machine learning are: voice recognition, image processing, and artificially intelligent vehicles. These applications of ML could not be built with traditional programming because the model(s) would see all of the data meant to build the app. In a classification problem, for example, a model is only exposed and recognizes patterns, based on a subset of the data also known as ‘the training data”. Likewise, a subset of remainder, is designated as “the test data” - to validate each party in a generalized global society.

Observation, Feature., Quantitative data vs Qualitative data.

In machine learning the term “observation” refers to a specific row within a data set. The name “feature” refers to the column values relative to the observations.There different kinds of data. Quantitative data, sometimes referred to as continuous, concerns floating point numbers. In some cases, quantitative data is called *numerical data*. On the other hand, information which describes a category, review, or event, is known as categorical data.

My interest in ML?

I am interested in machine learning. I have worked with it before in a variety of settings. I am looking forward to learning more about the statistical concepts involved with the field, as well as considering the business’ side of operations. I want to learn about how to handle data that is messy, such as real world data.