

A Definition of ML

Machine learning (ML) is the field of training computers to accomplish specific tasks by analyzing data and recognizing patterns in it.

Data, Pattern Recognition, and Accuracy

Machine learning is highly useful when the steps or rules to solving a problem are unknown to the programmer, or when the amount of data to be processed is high. Where traditional programming takes a deterministic approach to solving problems, where a known set of steps or rules are automated to process data and output the desired results, machine learning trains algorithms to recognize patterns in data and output predictions of new results. Machine learning algorithms require large amounts of data in order to produce more mathematically accurate models of the results. Determining the accuracy goal of a machine learning algorithm is the programmer's job -- ML algorithms need to meet a certain threshold for accuracy in order to be effective.

The Relationship Between ML and AI

Machine learning is a very broad and diverse field, with significant overlap with other related mathematical and computational fields: computer science, statistics, probability, and AI. Machine learning is a form of AI, along with other related fields like natural language processing.

Modern ML Applications

In modern times, we encounter ML often in our daily lives. One example is computer vision, in which our computers and smartphones are able to recognize our faces and fingerprints in lieu of passwords or pin codes. Traditional programming isn't capable of recognizing faces or fingerprints because, unlike the rules to match text input to a password or pin code, the rules to recognize these complex human features are less defined. Instead of implementing some rules to achieve this, a computer vision algorithm is trained on facial and fingerprint data to be able to recognize the patterns in people's faces and fingers. Another example is data analysis, which is used for tasks like traffic and weather prediction. These spontaneous and complex phenomena would be extremely difficult to encode in a set of rules, but a data analysis algorithm can be trained to recognize patterns in large amounts of traffic and weather data, and can thus predict when, where, and how these phenomena will occur.

Terminology

The data used for machine learning is normally organized into data frames, which are essentially tables. Data frames are described by the following terminology: "observation", "feature", "quantitative data", and "qualitative data". An observation is another name for a row, which represents a sample data point. A feature is another name for a column, which represent either quantitative or qualitative data describing each sample data point. Quantitative data is strictly numeric, meaning it can be any of an infinite set of numeric values. Qualitative data (also called factors) is categoric, meaning it can only be from a finite set of values (or "levels"). The organization of input data into rows and columns is important for analyzing any potential relationships between vectors, and the distinction between quantitative and qualitative data is important so that vectors of each type are treated in accordance to their type.

About My Interest in ML

My interest in ML coincides greatly with my interest in NLP. I realized that, through ML, I could infuse my programming skills with my lifelong passion for language, and so chose to pursue it. Besides NLP, I am also interested in applications like AI art because, as an artist by hobby as well, I am intrigued by both the possibilities as well as ethical concerns that innovations like this pose for human creators. My goal is to deepen my understanding of the ML field and perhaps work on projects that further human creativity, not diminish it.