Spencer Ng

At heart, machine learning (ML) is a process by which one trains computers to accurately recognize patterns in order to produce some sort of insight or result. To me, this implies that ML is founded on the idea of finding useful patterns that otherwise would be difficult to ascertain accurately, for one reason or another.

Within this process of ML are a few key requirements. Data is the avenue by which ML takes place, and as such represents the very foundation of all results of an ML approach. Because of this, acquiring clean, useful data is crucial for any project to succeed. Pattern recognition is the way in which algorithms generate insight into the aforementioned data. Without pattern recognition, data is a meaningless string of numbers, words, and events. It is pattern recognition that generates the story—and subsequent results—of ML. Finally, accuracy is the method in which ML techniques are measured. Accuracy is paramount in any field concerned with experimentation and prediction, but arguably doubly so for ML, where the methods and results of algorithms may not be fully clear or easily understood. Determining the accuracy of ML results therefore allows for knowledge of how useful predictions may be, and whether further work needs to be done.

Though frequently amalgamated into one concept, artificial intelligence (AI) and ML are separate ideas. Broadly speaking, artificial intelligence is the process by which systems are developed that mimic and perform tasks done by humans. ML is a subfield of artificial intelligence which aims to accomplish this goal by teaching models to learn automatically. Simply stated, all ML is AI, but not all AI is ML.

Examples of modern ML abound, including fraud detection, trading algorithms, facial recognition, chess engines, and far more. As a whole, ML is necessary because traditional programming techniques lack the ability to synthesize knowledge that is not already known. Because of this, for example, the best chess engine would be limited to the skill of the top human players. Additionally, ML thrives in situations of massive scales, which allows it to find patterns in data that humans simply could not comb through.

Key terms in ML include observations, features, qualitative data, and quantitative data. Observations are individual points of data, collectively which make up the dataset. Features are properties of observations—for example, a student may have a favorite class, a GPA, and an ID number. Qualitative data are defined by a limited number of values, such as the favorite class listed above, whereas quantitative data has a quantity, meaning that they are numeric.

Personally, ML is interesting because of the wide variety of use cases it has. As someone who has always loved unraveling stories and insights from data, ML seems to be a prime opportunity to do so at a scale that would previously have been impossible. In this class, I’d like to gain a working understanding of ML techniques and the ability to apply them to create things that are interesting and worthwhile.