Dillon Carter

Dwc180002

Overview of ML

Until around the 19th century, people really didn't have much knowledge of medical issues. There were guesses of what was wrong and what could be done to fix it, but these were largely ineffective. As the field of medical science advanced, doctors became better able to predict issues with their patients and assign better medical treatments. These weren't always accurate; there were often wrong diagnoses prescribed. Recently, machine learning algorithms fitted to patient data have started to be used for diagnosis. With the right model, the diagnosis returned could be accurate at a similar rate to highly experienced general practitioners [1]. The difference between one diagnosis and another could be a fractional difference between patients. This would be hard to detect with a traditional algorithm, requiring programmers to become extremely familiar with all different cases of how diseases can present. So, models can be used to analyze the data for the researchers and produce predictions where these edge cases are taken into account.

Machine learning is a field of study by which the data produced in the real world can be analyzed to find patterns using traditional mathematics combined with the speed and reliability of computers. By properly proportioning the data, statistical analysis and more complex mathematical processes can be run quickly to return information that ranges from a broad overview of the data to specific predictions.

The algorithms can be powerful, but it requires first the cleaning and processing of data beforehand. A myriad of inputs relating to all sorts of different concepts will not provide an accurate set to analyze and rather muddy up the output. If the size to age correlation is wanted, then it is best to provide it with a set of data composed of primarily the same thing, like of breeds of cats, rather than of a set of tropical fruits, vegetables, and wildlife. A pattern might be able to be detected out of the random set, but it likely won't be accurate. The data also should be ethically sourced. With the example demonstrated above, if the patients' records were given without consent, than the information returned cannot truly be used ethically.

Unethically sourced data could be demonstrated through the recent advances in machine learning powered art generators, of the like of Stable Diffusion. It was almost inconceivable a few years ago to think that a computer would be able to generate anything that looked as good as if a human had illustrated it. To create a program to mimic how a human illustrates would require creating a general process for art. Yet, with a large enough data set, a neural network is able to produce novel images from text. It finds patterns between pixels and their accompanying text that it can then use with a seeded noise image to produce something similar. However, the issue with these products arises in how they were trained. Copyrighted images from all different sources were used in the training of the models. Something that would have been clear cut in the past about whether the product was infringing on copyright is now slightly ambiguous.

To clean data, there are a few terms that data scientists use to categorize it. Information about the real world will come in either qualitative or quantitative forms. Qualitative data is information that can belong to a set of data. It can take on only 1 value from the set. Quantitative data is information that is numeric. Each data point is an observation which can have multiple features relating to the different measurements taken for the data point. Patterns can be drawn by looking at features against other features, which labels them as predictors.

My interest coming into machine learning comes from wanting to apply it to games. There is a YouTube channel which I watch, CodeBullet, that focuses on strange projects that use machine learning to implement AI behaviors in different games and simulations he codes. That was the initial spark for my interest. Another channel that got me more interested in the math side was Stand-Up Maths, where he would demonstrate the application of machine learning in, for example, one video classifying whales. My two goals for this class are to 1) satisfy my curiosity with the subject and 2) discover ways that I can potentially use what I learn for games. I think its immensely interesting how the usage of patterns can produce behavior that seems intelligent and would love to understand how that can be implemented into either creating more realistic behavioral simulations or assisting in providing accessibility to people without creating a much larger burden on the developers.

References:

[1] https://www.nature.com/articles/s41467-020-17419-7#Sec3