

Sunni Magan

CS 4375.004

January 28, 2023

Overview of Machine Learning

Machine learning is the process of using computer programming in conjunction with statistical models to make inferences on data in order to make the data more insightful and useful.

Machine learning can be implemented in virtually any field provided there is a sufficiently sized data set. A large amount of data is one of the most important aspects of machine learning, without which it would be difficult to create any useful models. However, the size of the data is not the only prerequisite for good quality data. In order to create an accurate model, the data must also be representative of the population it samples from. Otherwise, any pattern drawn by the model would not generalize well for the population. The amazing thing about machine learning is, that given enough data, it can find many patterns in the data that would go unnoticed by a human which makes machine learning an invaluable tool in any industry. However, the conclusions drawn by the machine learning model will not be perfect. The model is only making a prediction based on known data, therefore there will be a certain percent of error in the model. While unavoidable, it is possible to tune the learning model to lower the percent error. However there is a point of diminishing returns and a well trained model could have high (+99%) accuracy.

A popular field which largely implements machine learning is Artificial Intelligence. Machine learning and AI often get mislabeled as the same thing. While they are closely related, one is actually only a small part of the larger picture. Artificial Intelligence is what the name implies. It is an attempt to create a computer that resembles the thought and decision making process of humans [1]. AI achieves this by using advanced machine learning models, known as deep learning.

Netflix, while not advertised, is a perfect example of the use of basic machine learning principles. They are most well known for their user recommendation technique[2]. There are many ways of doing this, but one way would be to create some sort of correlation between

movies the user has seen and liked and movies that the user has not seen. The higher the correlation of movies, the more likely a user would like this new movie. This type of algorithm would not work in a traditional programming environment because in order to compare movies, it would have to be done iteratively for every new movie. This would take too long considering the size Netflix's movie database.

Another example of a well known use of machine learning would be any one of the many virtual assistants (Siri, Alexa, Google Assistant). These systems use a variety of machine learning principles to the point where they verge on being considered "AI". They all deploy a technique called Natural Language Processing to convert spoken and written word into machine usable data. This would be very difficult to implement in a traditional programming environment. Because the computer has no context of the input, word or command would need a brute force solution. This is impractical in relation to language, because it would take years to list every single possible command, let alone processing them. Machine learning is able to build a language context based on rules or tokenization in order to parse a language on the fly [3].

As stated earlier, one of the prerequisites of machine learning is to have a large amount of data. In order to make handling the data easier, terms were created to better identify each component of data. Most data is a collection of data points. Each data point is known as an observations. Each observation also has a collection of attributes called features, These features can either be quantitative or qualitative. A quantitative attribute is one that can be represented using discrete or continuous numbers. Qualitative data is used to describe or categorize the data. This break down of the data is important because it helps us organize the data with a process called feature engineering to allow the data to be used in multiple ways to solve different problems.

Machine learning interests me because of the limitless opportunities and fields it can be used in. Certain fields like cybersecurity, while every company needs it, would have you do the same thing wherever you go. Where as machine learning will serve a different purpose in any environment. Machine learning can also tell us information about the data and ourselves that we would not otherwise realize.

Works Cited

- “*Artificial Intelligence (AI) vs. Machine Learning*,” CU-CAI, 03-Mar-2022. [Online]. Available: <https://ai.engineering.columbia.edu/ai-vs-machine-learning/>. [Accessed: 28-Jan-2023].
- “*Netflix Research*” Netflix Research. [Online]. Available: <https://research.netflix.com/research-area/machine-learning-platform>. [Accessed: 28-Jan-2023].
- B. Lutkevich and E. Burns, “What is natural language processing? an introduction to NLP,”* Enterprise AI, 20-Jan-2023. [Online]. Available: <https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP>. [Accessed: 28-Jan-2023]. *K. Mazidi, Machine Learning Handbook.*