Dylan Kapustka

Dr. Mazidi

CS 4375.003

September 4, 2022

Overview of ML

Machine Learning (ML) is the process in which a computer learns from data, adapts, and improves without explicitly programmed instruction.

Data plays a crucial role in ML as it is how models are trained. Without data, nothing can be determined, and ML is not possible. Taking many forms, data allows for training, validating, and testing on ML systems, leading to accurate predictions. Pattern recognition in data allows algorithms to learn and determine outcomes. For example, credit risk assessment, relationship compatibility, etc., can all be determined by patterns in data. Further, these patterns allow for systems to predict outcomes on future data that the algorithm has yet to process. When evaluating a trained model, we want to see how well the ML system can perform on unseen data, leading to accuracy. Predictions need to be accurate as possible, and one must ensure that the system is not just making random guesses. To ensure this, ML utilizes multiple measurement techniques that determine accuracy and algorithm performance. These measurements can later be used to see what changes, if any, need to be made to the training model.

Artificial Intelligence (AI) and ML are very closely connected but are not the same. AI is where a computer system can mimic that of human cognizance when it comes to learning and problem solving, and ML helps a computer learn without explicitly programmed instruction, using models of data. ML is a subset of AI. When looking at AI, one will think of an intelligent computer that can think like a human. ML is how that computer developed its intelligence.

Image Recognition is one example of a modern ML application, where traditional programming is not viable. For example, say we wanted to know if a certain image contained a dog or cat. In traditional programming, one would have to write hundreds of conditional statements for determining just a single image. Further, if that image were to change, the entire program would break. We can avoid this by using an ML system that has processed thousands of image data points. Language translation is another example of a modern ML application. In language translation utilizing ML, the system can handle thousands of languages, pull from dictionaries, utilize NLP for accurate translation, perform POS Tagging, Named Entity Recognition, and more. Meanwhile a traditional programming language would require tedious work just to handle a single sentence.

In ML, an **observation** is a data point, row, or sample in a dataset. This is important, because as stated earlier, without data nothing can be done. An **attribute** could be seen as a column header or quality that describes an observation such as (GPA, Hours, SAT, etc.). This is important because it is used to build, test, and score a learning model. **Quantitative data** is numerical data that has been structured to rely on measurement and objectiveness. It is important

for training ML algorithms because the numerical values provided can be analyzed statistically. **Qualitative data** is non-numerical data (text, images, etc.). The benefits of qualitative data are that it can improve accuracy through providing additional context and can also allow for smaller training datasets.

Personally, I have always had a fascination with ML, and see myself working with it in the future. As far as avenues for learning go, I look forward to this course and think that many of the class projects will be a great step forward in my understanding. Expanding beyond the scope of the class, I may take some courses online that dive deeper into specified categories of ML, such as image recognition and more, so I can apply them to personal projects and leverage that skillset for professional jobs. Overall, I look forward to this course and what we will be covering in it, and hope to come out this semester, with a much stronger understanding of ML.