**Huy Dai Nguyen**

**Class Project**

**MACHINE LEARNING**

Machine learning is “a method of data analysis that automates analytical model building." It’s a branch of Artificial Intelligence based on systems which, with minimal human intervention, can analyze data, identify patterns and make accurate decisions. There are 3 types of machine learning: supervised, unsupervised and reinforcement learning. In supervised learning, the process is similar to a student learning from his teacher. Example data and associated target responses are inputted into the algorithm, allowing it to predict the correct response when posed with new similar examples. Common algorithms include decision trees, linear regression and neural networks. An example of supervised learning would be to train a model to recognize and categorize new photos based on previous ones with information included. In unsupervised learning, algorithm learns from plain examples without any associated response. In other words, there is no teacher involved because he/she doesn’t know what to look for in the data. Through the examples, the algorithm can determine patterns in data and teach human expert things he/she isn’t aware of. Common algorithms include k-means clustering and association rules. An example would be categorizing a set of photos based on individuals through face detection. Finally, in reinforcement learning, the algorithm is presented with unlabeled examples that are accompanied with positive or negative feedback according to the solutions it proposes. In this case, the A.I is learning through trial-and-error. Such concept is implemented into Google’s AlphaGo where computer analyzes pixel on a screen to teach a neural network to play a Chinese game called Go, beating the world master in the process.

Machine learning has been around for a long time. One of earliest invention of machine learning is the Turing Test by Alan Turing in 1950. The Turing Test was used to determine if machine has real intelligence. Results are evaluated based on whether a person knew if it was a machine or a person they were talking to. However, the hype surrounding ML on news media didn’t begin until the past 5-10 years, starting with the success of IBM’s Watson on the TV game show Jeopardy in 2011. There are a few explanations as to why this has happened. Some say that with the improvement of technology, algorithms are further developed, bringing error rate from 0.8% in 1998 down to as low as 0.23% in 2016. But more importantly, the advent of cloud computing has made machine learning much cheaper and easier. Through cloud computing services from Amazon, Microsoft, Google etc., developers can run massively large computing jobs without having to pay up-front and computer maintenance cost. This led to the change from supercomputers, which were as large as a room and cost a lot to manage, to laptops and portable devices. Nowadays, funding for machine learning has never been more affordable. Funding and technological innovation have paved way for Deep Learning, creating unsupervised and reinforcement learning machine learning.

Some people think machine learning is a more glorified version of statistics. They believe that in the future, machine learning will render statistics obsolete. However, such idea is untrue. While statistics and machine learning both focus on mathematical techniques and use data to solve problems, they have different goals and strategies. Statistics concerns primarily with model validity, parameter accuracy and assumptions (linear regression, logistics regression) to explain causation. Machine learning, on the other hand, works to generate the best prediction method so that the decision is automated. It doesn’t have to worry about model assumptions or diagnostics, but it cannot explain causation. In short, statistics is used for scientific research and machine learning is used to design real-world algorithm.