**Machine Learning**

**What is Machine Learning?**

**Arthur Samuel**, a pioneer in the field of artificial intelligence and computer gaming, coined the term **“Machine Learning”**. He defined machine learning as – **“Field of study that gives computers the capability to learn without being explicitly programmed”**. In a very layman manner, Machine Learning (ML) can be explained as automating and improving the learning process of computers based on their experiences without being actually programmed i.e. without any human assistance. The process starts with feeding good quality data and then training our machines (computers) by building machine learning models using the data and different algorithms. The choice of algorithms depends on what type of data do we have and what kind of task we are trying to automate.[[1]](#endnote-1)

**Using a linear regression model**

The term linear model implies that the model is specified as a linear combination of features. Based on training data, the learning process computes one weight for each feature to form a model that can predict or estimate the target value. [[2]](#endnote-2)

**Resources**

## Data Normalization

Using data collected through the Census API, which was saved in the database, we were able to create a linear regression model in python with the use of keras. However, prior to the used of the data in our model we had to first normalize the input. Through the use of pandas we selected very specific data to be input. Then, using the sklearn library we were able to normalize our data to a number between 0 and 1 to help improve the prediction made by the model. The scalars used were then saved to unnormalize the prediction and obtain a more user friendly output.

## Keras w/ TensorFlow

Using the normalized census data from the year 2017. A model was created using keras backed with TensorFlow as we found more visualization options when using this format. (Theano is another library you can use to back keras for machine learning.) The model was then trained on random sample of 70% of data split into 2 sections. One contained the various attributes while the other obtained the targets (this is the value you wish to have the computer predict.) The remaining 30% of the data was then split into 2 sections as well same as above but was used for the computer to test its accuracy. At varying points we used varying numbers of layers, optimizers, and batch sizes. Each change in these settings creates a change in the weights applied to the attributes, certain changes will improve the accuracy of your model.

## Accuracy

## As the model is training the accuracy of its guess of the test data targets is compared to the actual test data target providing an evaluation of the accuracy of the guess. In our model we mainly relied on the mean squared error value which was calculated using the keras library evaluate function.

## Prediction

## Once the model has been created to your satisfaction and preferably saved. It can then be loaded or used to create a prediction for data it has never seen. To obtain an accurate prediction the data for all the attributes it was trained on should be normalized in the same fashion and provided.

1. <https://www.geeksforgeeks.org/ml-machine-learning/> [↑](#endnote-ref-1)
2. <https://docs.aws.amazon.com/machine-learning/latest/dg/linear-models.html> [↑](#endnote-ref-2)