# MACHINE LEARNING HANDEDNESS

CIS 481: Computational Learning

Castellanos, Alexis
acastel@umich.edu

Alexis Castellanos
Dr. Luis Ortiz
April 24th, 2020

# Machine Learning and Handedness

## Background

Dating back to the Paleolithic era, researchers have found that approximately 10% of humans are left-handed. Interestingly this percentage of left-handed people in the world has remained timely the same. There is no agreed upon explanation as to why the percentage has remained the same. It is astonishing that Handedness is the product of genetics, probability and environmental factors and has remained the same generationally (Dove, 2015).

Identifying handedness based on handwriting has many implications for fields such as Forensics Science, Psychology, Biometric Security and Graphology (Morera, Sánchez, Vélez & Moreno, 2018). Forensic Scientists could use this model to analyze text to distinguish criminals' handedness and narrow down investigations. Computational machine learning of handedness can be useful for Psychology in order to further the current research on handwriting analysis. It has the potential to build stronger correlations between handwriting and personality attributes. Along that notation it would be beneficial to the field of graphologists. Graphologists often struggle identifying handedness due to high variety in handwriting styles Biometric Security to find benefits in a handedness model in order to improve computer systems security (Morera, Sánchez, Vélez & Moreno, 2018).

## Introduction

Hand dominance interests me because I am left-handed along with all members of my immediate family. The specific aim of the project is to determine between left-handed and right-handed handwriting. The complete total of the dataset was 400 text samples. 260 samples were used to train the model with a breakdown of 195 right-handed text samples and 65 left-handed samples. 70 samples were used as validation data with a breakdown of 58 right-handed and 22 left-handed samples. The remaining 70 samples were used for testing and had the same distribution of right and left-handed distribution. Two models were built that use two different metrics, one mean square error and the other uses binary cross entropy for evaluation. I decided to use both metrics in order to compare the two models to one another. This comparison strengthened my understanding of neural networks.

## Design Choices

I have chosen to create a CNN as my architecture for this problem. I was inspired by natural vision perception because convolutional neural networks are often used in visual problems. For this reason, the architecture I created to solve this problem was sequential (standard) consisting of four convolutional and max pooling layers followed by a flatten layer and two dense layers.

A feature I couldn't change was what was considered to be the success rate of model runs due to the imbalance of handedness. 90% of the human population is right hand dominant. Therefore, for my model to be considered successful it should have above 90% accuracy.

Continuing off the problem of imbalanced data, considering data handedness distribution I chose instead of doubling left-handed class to use fit generator to fit the model. I found that this technique saved me time and worked efficiently with the dataset.

**Approach**

Data: The sample texts used to build the model were gathered from people at the University of Michigan Ann Arbor and Dearborn campuses. The participants were instructed to state their hand preference (left and right) and to use their dominant hand to write a message (their name was suggested to write). A total of 400 clean samples were collected of them 109 were left-handed text samples and 291 right-handed text samples.
The data was split up into three labeled sets each containing two sets:
Training (total 260, 195 right handed 65 left ) , testing ( 70 : 58 right , 22 left ) and validation ( 70: 58 right 22 left ).

Data Pre- Processing: In order to build the model, I first needed to understand how to read picture files. Once I decided the JPEG content to RGB grids of pixels, I was able to move into building the model. Next I converted these into floating point tensors and rescaled the pixel values between 0 - 255 to the [0,1] interval. Finally, I resize all images to 150x150.

Architecture: Sequential: The model has four Convolutional layers with relu activation. Each one is followed by one max pooling layer. When the four layers pass, a flatten layer is used. Next a dense layer with relu activation is applied. Finally, a dense layer with sigmoid activation becomes in use.

Compiling: I completed two different scripts: mean square error and binary cross entropy, in order to optimize the accuracy. Another technique I used was to improve callbacks and early stopping by saving the best model and storing the best weights found during training.

Fitting: I applied the fit generator to our training data through 50 epochs against our validation set. We next saved the model and re-ran the process again. Interestingly, 100 epochs were passing through in our best model as a parameter.

**Results_**

The final results for the Mean Square Error Loss score was 0.0380 with an accuracy score of 0.9571. Meanwhile the Binary Cross Entropy Loss reached a 0.9571 accuracy score and loss score of 0.1199. Interestingly despite the binary cross entropy resulting in the same score as the mean square error loss the binary cross entropy had much higher loss. However, I still believe the binary cross entropy score to be a more accurate representation of the performance because it's standard to use binary classification for this type of problem. Furthermore, I believe to have run into hardware limitations to train my model on my last data set, Test.

**Limitations**

COVID-19 was at the epicenter of my limitations causing data collection, time, and model constraints. The pandemic caused a fragmented workflow making it challenging for me to increase my data set. A small data set affects the models score results. Furthermore, it was difficult to implement more model improvements. Limited peer reviewed literature was also a large obstacle. Having limited literature to build off of made starting the project difficult.

**Conclusion**

The project aims to create a machine learning model with the ability to classify a person by handedness, based on the individual's handwriting. Overall, I believe a binary cross entropy is the best model to use. I gained strong experiences by building two models and running them. The project has many fields that could benefit from further findings on handedness.

## References

Dove, L. (2015). There's a Study on That: The Stone Age Had Just as Many Lefties. Retrieved 24 April 2020, from https://science.howstuffworks.com/life/inside-the-mind/human-brain/theres-a-study-that-the-stone-age-had-just-many-lefties.htm

Morera, Á., Sánchez, Á., Vélez, J., & Moreno, A. (2018). Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks. *Complexity*, *2018*, 1-14. doi: 10.1155/2018/3891624