HANDWRITING RECOGNITION

Introduction Attempt One

Once upon a time, when you were very young -- not even a seedling in my tummy yet! -- your father and I would use this long, thin tool, called a pen, yes PEN. Can you spell that for mom? P - E - N. Exactly. We would use this *pen* to write -- yes, this is before writing on the computer -- we would write notes to one another. No, no not on your tablet, the pen was used with paper. Oh sweet girl, no, not like the paper from the toilet, no. Well, I guess kind of like the paper from the toilet? Just imagine a little sturdier. And about the same size as dad's big tablet. So yes, we would write one another little letters by hand, using pen and paper. It was really quite romantic. Unfortunately many of them got lost during the big fire... what was that? Oh, yes. The big fire that burnt mom and dad's first home. And

grandma and grandpa's. Great memory! But yes, it was very sad. We used to live -- Honey can you grab that old map from your office? Yes the framed one. -- much farther from the ocean because the ocean used to be all the way out here. See this? That used to be the coast! Before the fires and the flooding, this was one of the most beautiful places in the whole world. I agree that it's still very beautiful, but you, little one, you are what makes it beautiful -- right Honey?!

Thankfully, your mom has always been a digital hoarder, so she insisted on digitizing those letters forever ago. Let's see if I can pull some up right now. I might not be able to read my writing! Honey do you remember cursive?! One of my most favorite days in elementary school -- yes, elementary school is similar to your primers, yes, but imagine going to a big building, a big SQUARE building with all the other kids. Not virtual school, like, I could reach out and touch my seat partner. Just like I'm touching you right now! Right Honey? Oh man, I miss the smell of those sticky classrooms -- and library books!! What's a library book? You know what, let's bookmark that for now. That just means to save our spot so we can return to it later. Yes, they are all very related. Dad will forever be sad he didn't get to show you a real book. No, no. Not a book on the tablet. Like, a book book. With paper. Yes, like paper from the toilet.

Introduction Attempt Two

THE YEAR IS 2074, BOOKS AND PAPERS ARE SACRED AND HOUSED ONLY IN MUSEUMS.

In the early days of humanity, one of the ways society would communicate, both individually and as a group, was through writing. Before the printing press, everything was handwritten. NOTE: In this context handwritten doesn't mean typed-by-hand. It means literally handwritten. Again, **not typed**. Written. Even after the invention of the printing press -- and even for a while after the invention of the typewriter, computer and holotab -- handwriting was still something that humans did. As technology increased, handwriting became more and more niche, used almost exclusively for art or gonzo historical analysis.

Analysis & Models

THE DATA (FROM KAGGLE)

The training data set, (train.csv), has 785 columns. The first column, called "label", is the digit that was drawn by the user. The rest of the columns contain the pixel-values of the associated image.

Each pixel column in the training set has a name like pixelx, where x is an integer between 0 and 783, inclusive. To locate this pixel on the image, suppose that we have decomposed x as x = i * 28 + j, where i and j are integers between 0 and 27, inclusive. Then pixelx is located on row i and column j of a 28 x 28 matrix, (indexing by zero).

These libraries were used: e1071, dyplr, naivebayes. The data was downloaded from kaggle.

FIRST ATTEMPT

- The first column was removed
- The data was reduced in size
- The data was split into test and train
- The data was analyzed

SECOND ATTEMPT

THIRD ATTEMPT

FINAL ATTEMPT

Results

Conclusion

NOTES

- Handwriting may seem old fashioned / antiquated, but it's still useful

- Handwriting can tell us how language develops over time
- Evidence that humans learn better when they physically write
- Handwriting analysis can help us digitize old documents
- Forensic handwriting analysis can help catch criminals like Robert Durst and the BTK killer
- Analyzing handwriting can be difficult because everyone's handwriting is different. More than one method
- **HANDWRITING HERE TO STAY:**
  - Handwriting is still` used for everything from note taking to love letters to grocery lists
  - Evidence that humans learn better when they physically write
  - Handwriting is still used widely in lesser developed nations
  - Handwriting is still used widely in schools, offices and homes
- **HANDWRITING ANALYSIS HELPS US:**
  - Handwriting analysis can help us digitize old documents
  - Handwriting can help us learn how language develops over time
  - Forensic handwriting analysis can help catch criminals like Robert Durst and the BTK killer

RCODE

```
## =================================
## STEP 1: IMPORT LIBRARIES
## =================================
library(e1071)
library(naivebayes)
library(dplyr)


## =================================
## STEP 2: IMPORT DATA
## =================================
trainNB <- read.csv("/Users/kosburn/syracuse/IST707/WK7/handwriting_train.csv", header = TRUE)
testNB <- read.csv("/Users/kosburn/syracuse/IST707/WK7/handwriting_test.csv", header = TRUE)

## 2a. Working backup
originalTrainNB <- trainNB
originalTestNB <-testNB
```

```
## ===================================
## STEP 3: TRANSFORM DATA
## ===================================

## 3a. Examine the data
head(trainNB)
str(trainNB)
head(testNB)
str(testNB)


## 3b. Turn ints to num
## IS THIS THE BEST WAY OMG HELP
## -- FOR TRAINING DATA
trainNB <- trainNB %>%
  mutate_all(as.numeric)
## -- FOR TEST DATA
testNB <- testNB %>%
  mutate_all(as.numeric)


## 3c. REMOVE THE LABELS FIRST KENDRA DUH
## -- FOR TRAINING DATA
trainNB_labels <- as.factor(trainNB$label)
trainNB_nolabels <- trainNB[-c(1)]
## -- FOR TEST DATA
## Test data doesn't have labels, obviously

## 3d. Convert anything > 0 to 1
## Either pixel has writing or doesn't have writing
## -- FOR TRAINING DATA
trainNB_nolabels[trainNB_nolabels > 0] <- 1
## -- FOR TEST DATA
testNB[testNB > 0] <- 1

## 3d. Re-assemble TRAINING data
## _T means "transformed"
trainNB_T <- data.frame("label" = trainNB_labels, trainNB_nolabels)

## 3e. Turn EVERYTHING INTO A FACTOR
## -- FOR TRAINING DATA
trainNB_T <- trainNB_T %>%
  mutate_all(as.factor)
## -- FOR TEST DATA
testNB <- testNB %>%
  mutate_all(as.factor)


## ===================================
## STEP 4: CREATE THE MODEL
## ===================================

NB_e1071<-naiveBayes(label~., data=trainNB_T, na.action = na.pass)
NB_e1071_Pred <- predict(NB_e1071, testNB)
NB_e1071
```

```
table(NB_e1071_Pred)
NB_e1071_Pred
```