

# Number Recognition using R

Justin Heck

justinheck@u.boisestate.edu

Monte Hedrick

montehedrick@u.boisestate.edu

David Marcial

davidmarcial@u.boisestate.edu

Donovan Wright

donovanwright@u.boisestate.edu

October 2020

## **1 Abstract**

For this project we plan to develop a machine learning model that is able to successfully classify numbers 0-9. We plan to train the model on the MNIST data set using the TensorFlow for R library. The MNIST data set will consist of 60,000 images for training and 10,000 images for testing. Once we have created a working model that is able to classify the numbers 0-9, we would like to expand the model and have it classify mathematical operands such as plus, minus, multiply, and divide. This machine learning model in the end will only serve as a learning purpose for our team to expand our knowledge of R, along with our machine learning knowledge.

## **2 Introduction**

R is considered to be a language that is centered around data science. Our group wants to explore the use of R in a machine learning application. Since we do not collectively have much experience with this topic, we decided to go with a relatively simple project idea. That is recognizing hand written numbers. If we are able to achieve this, our goal to stretch ourselves will be to apply that in recognizing simple mathematical equations and evaluating them.

## **3 Project Description**

### **3.1 Base Goal**

Our base goal for this project is for our machine learning application to be able to recognize the handwritten numbers 0-9. We plan on achieving this by using the TensorFlow for R library provided by google. By using this library we will train and test our model on images provided by the MNIST dataset[1].

Example MNIST data



Figure 1: A series of examples MNIST images[2]

### 3.2 Long Term Goal

Ideally our long term goal would be for the application to recognize the handwritten number and perform basic Math operations such as addition, subtraction, multiplication, and division.

## 4 Evaluation

We will have several different success criteria for our project. With the MNIST data-set of handwritten numbers, there are specific testing data-set that we will use to verify that our program is correctly identifying each input. We want to get to the point where we have a 90% accuracy rate in identifying the numbers. As a part of our stretch goals, we would eventually like to get far enough along that we can run the program and have it recognize the handwriting of students in the class.

## 5 Management

### 5.1 Timeline

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6 - Onwards
Research and Planning	Process Data Set	Create Model	Model Testing	Finish Base Goal	Long Term Goal

## 5.2 Communication

We plan on using a Discord server to help our group communicate with each other. One way we will use it is that whenever we are working on the project we will join the voice chat so we can more help one another as we are working. We also plan on having a weekly meeting on Fridays at 4pm where we review what work was done during the week, help wrap up anything that was holding us up, and plan what needs to be accomplished the following the week

## 5.3 Division of labor

The division of labor will be decided each week during the Research and Planning phase of the project, and will be subsequently decided during each weekly meeting. Sections of the project may be worked on individually or jointly by members of the group. Work should be divided as evenly as possible amongst group members, and work assignments should be completed before each weekly meeting.

## References

- [1] Yann LeCun, Corinna Cortes, and Christopher J.C. Burges. "*The MNIST database of handwritten digits*". Yann LeCun's Website [yann.lecun.com](http://yann.lecun.com). Retrieved 13 October 2020.
- [2] Josef Steppan "*Sample images from MNIST test dataset*". Josef Stappan's wikipedia addition [en.-wikipedia.org/wiki/MNIST\\_database](https://en.wikipedia.org/wiki/MNIST_database). Retrieved 13 October 2020.