CSC3060 AIDA – Assignment 2

Thomas Pickup

40145342

11 January 2019

# Introduction

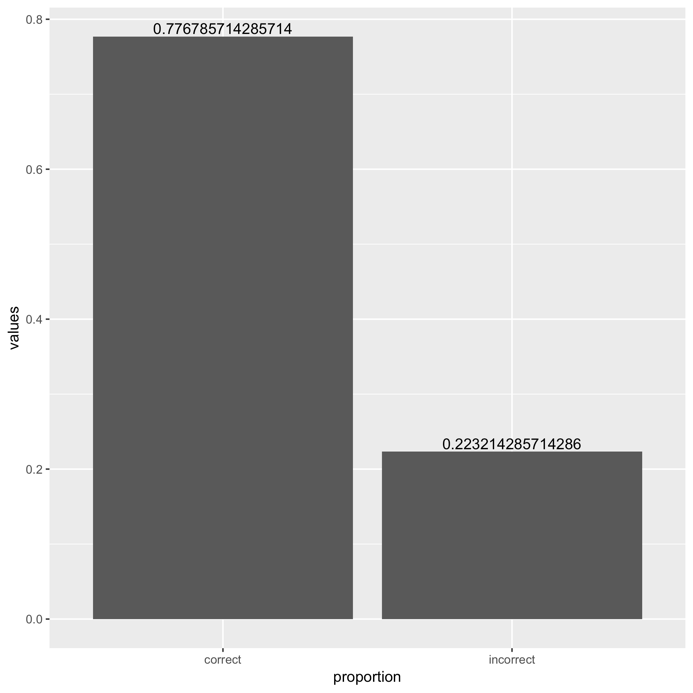
This report follows on from Assignment 1 for the Artificial Intelligence and Data Analytics module. In Section 1, I will build upon the features gathered from the handwritten images in Assignment 1 by using a Logistic Regression model to attempt to predict whether a character is a digit or a letter.

For the next three sections, I will use a dataset that was provided to me. I will explore three different classification tests to predict what a character is. The first of these, in section 2, will be the K Nearest Neighbours algorithm. I will evaluate the KNN Algorithm both with and without cross validation. The second algorithm is done by building a decision tree, this will be done in section 3 along with the random forest classification algorithm. Finally, in section 4, I will build a model that predicts the symbol for an unknown test item. This set of test items will be provided without the labels.

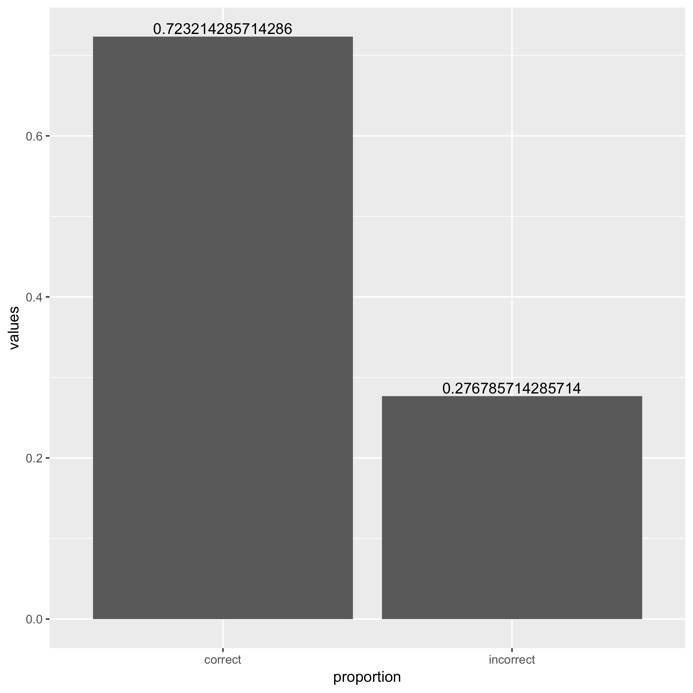
I will be using R for all of the code in this Assignment. After I have explored each of these sections, I will write up my conclusions in this report.

# Section 1

## Section 1.1



## Section 1.2



## Section 1.3

|  |  |
| --- | --- |
|  |  |
| Part 1 Accuracy Graph | Part 1 Accuracy Graph with Cross Validation |
|  |  |
|  |  |
| Part 2 Accuracy Graph | Part 2 Accuracy Graph with Cross Validation |

## Section 1.4

## Section 1.5

# Section 2

## Section 2.1

## Section 2.2

## Section 2.3

# Section 3

## Section 3.1

## Section 3.2

# Section 4

# Conclusions