CSC3060 AIDA – Assignment 2

Thomas Pickup

40145342

11 January 2019

Table of Contents

[Introduction 3](#_Toc534636608)

[Section 1 3](#_Toc534636609)

[Section 1.1 3](#_Toc534636610)

[Section 1.2 3](#_Toc534636611)

[Section 1.3 4](#_Toc534636612)

[Section 1.4 5](#_Toc534636613)

[Section 1.5 5](#_Toc534636614)

[Section 2 5](#_Toc534636615)

[Section 2.1 5](#_Toc534636616)

[Section 2.2 6](#_Toc534636617)

[Section 2.3 6](#_Toc534636618)

[Section 3 6](#_Toc534636619)

[Section 3.1 6](#_Toc534636620)

[Section 3.2 6](#_Toc534636621)

[Section 4 6](#_Toc534636622)

[Conclusions 6](#_Toc534636623)

# Introduction

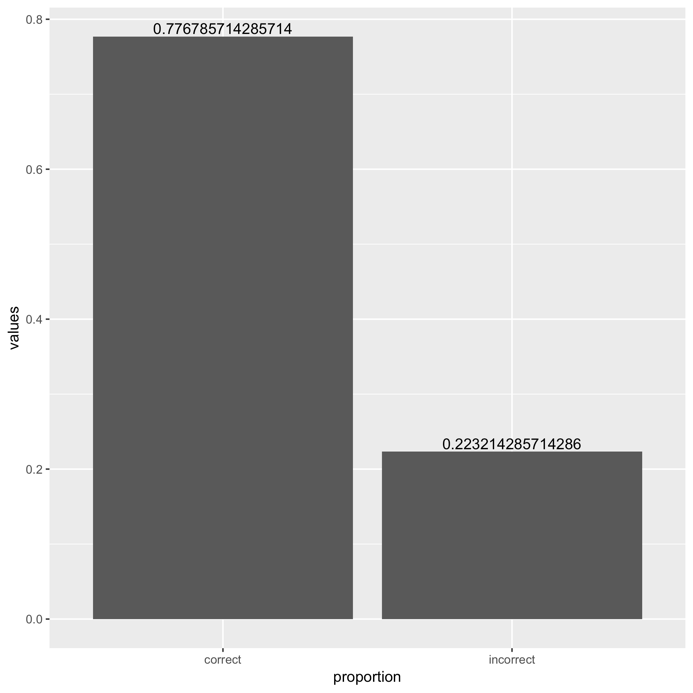
This report follows on from Assignment 1 for the Artificial Intelligence and Data Analytics module. In Section 1, I will build upon the features gathered from the handwritten images in Assignment 1 by using a Logistic Regression model to attempt to predict whether a character is a digit or a letter.

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.

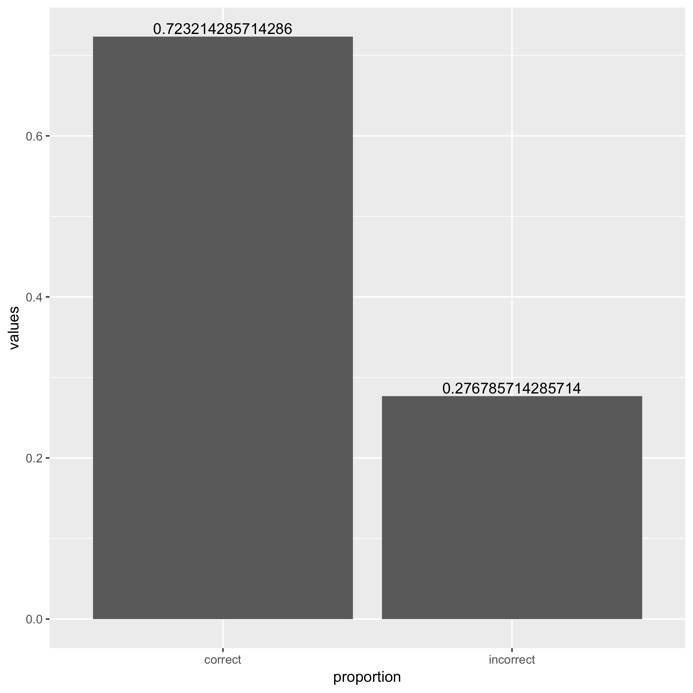
I will be using R for all of the code in this Assignment. After I have explored each of these sections, I will write up my conclusions in this report.

# 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 |

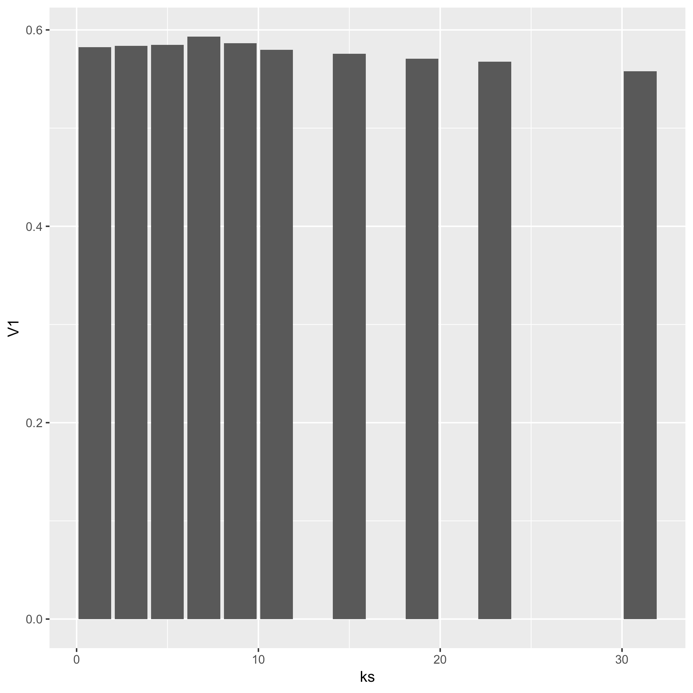
## Section 1.4

## Section 1.5

# Section 2

## Section 2.1

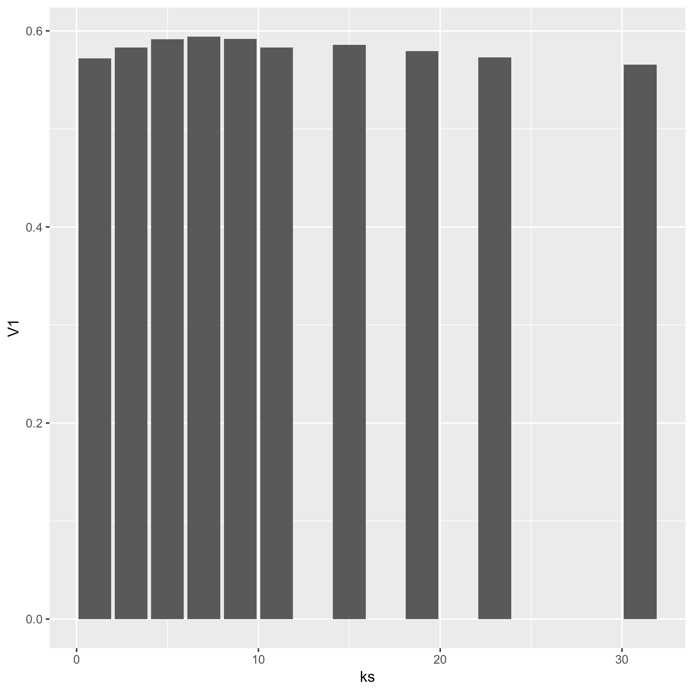
djjdjdjd



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **K Value** | 1 | 3 | 5 | 7 | 9 | 11 | 15 | 19 | 23 | 31 |
| **Accuracy** | 0.5822785 | 0.5837854 | 0.5846896 | 0.5931284 | 0.5864979 | 0.5795660 | 0.5756480 | 0.5705244 | 0.5675105 | 0.5578662 |

## Section 2.2

djdjdjd



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **K Value** | 1 | 3 | 5 | 7 | 9 | 11 | 15 | 19 | 23 | 31 |
| **Accuracy** | 0.5719955 | 0.5829932 | 0.5916100 | 0.5941043 | 0.5918367 | 0.5831066 | 0.5857143 | 0.5792517 | 0.5730159 | 0.5654195 |

## Section 2.3

# Section 3

## Section 3.1

## Section 3.2

# Section 4

# Conclusions