574 - MACHINE LEARNING

Programming Assignment 3

CLASSIFICATION AND REGRESSION

TEAM 28-

* Ranganath Sundar (50169271)
* Shalmalee Vishwasrao (50169915)
* Shriya Hardikar (50168340)

Introduction

In this assignment, the task of classifying hand-written digit images has been implemented using the methods

1. Logistic Regression
2. Support Vector Machine

Logistic Regression

Logistic Regression was applied to training, validation and testing data.

The observed accuracy using binary classifier is:

* Training set accuracy : 86.206%
* Validation set accuracy : 85.36%
* Test set accuracy : 85.44%

The observed accuracy using multi class regression is:

* Training set accuracy : 93.39%
* Validation set accuracy : 92.43%
* Test set accuracy : 92.67%

Support Vector Machines

The method of support vector machines was applied to the same data of hand-written digits. The observations are as below.

1. **Using Linear Kernel**

* Training set accuracy : 97.286%
* Validation set accuracy : 93.64%
* Test set accuracy : 93.78%

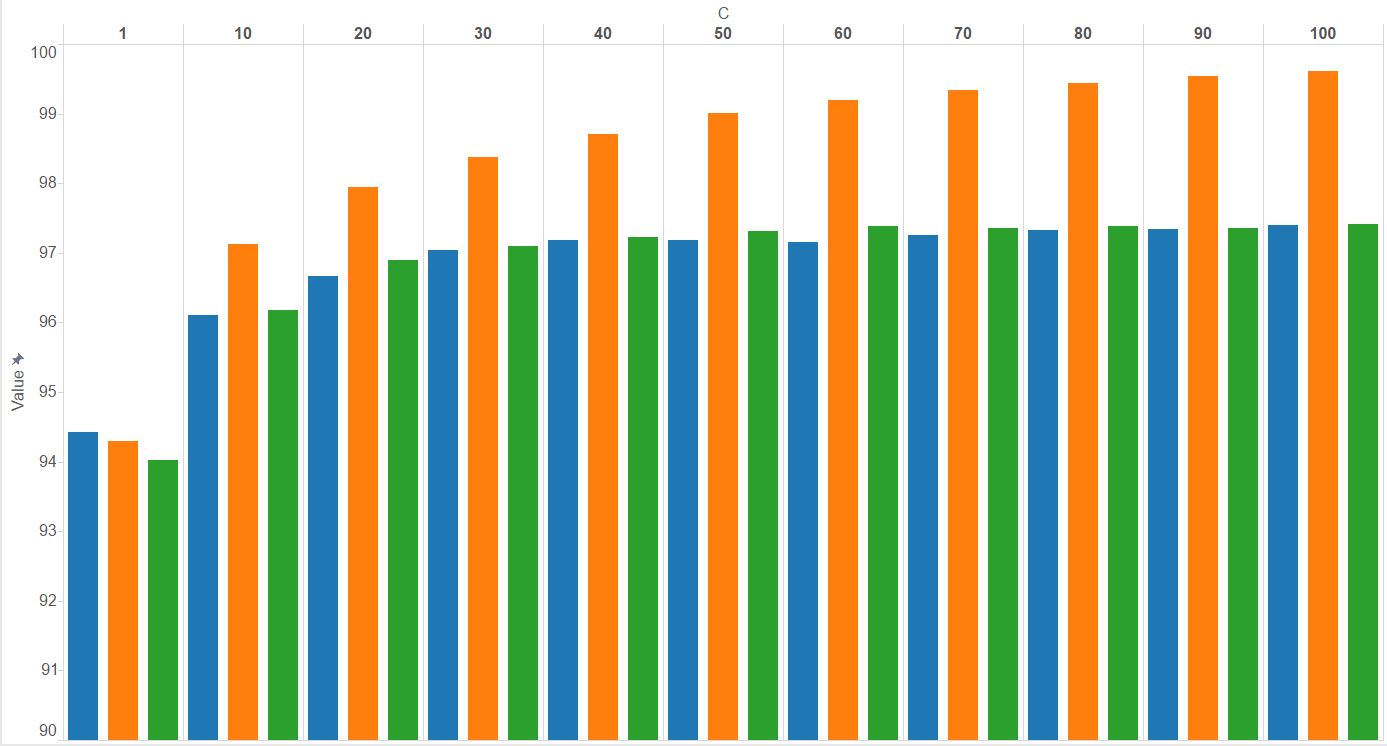
1. **Using Radial Basis Function with Gamma = 1**

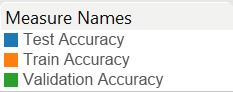
* Training set accuracy : 100%
* Validation set accuracy : 15.48%
* Test set accuracy : 17.14%

1. **Using Radial Basis Function with Gamma = 0.1 (default)**

* Training set accuracy : 94.294%
* Validation set accuracy : 94.02%
* Test set accuracy : 94.42%

1. **Using Radial Basis Function with Gamma = 0.1 (default) and varying the value of C** as follows:





|  |  |  |  |
| --- | --- | --- | --- |
| **C** | **Training set Accuracy** | **Validation set Accuracy** | **Train set Accuracy** |
| 1 | 94.294 | 94.02 | 94.42 |
| 10 | 97.132 | 96.18 | 96.1 |
| 20 | 97.952 | 96.9 | 96.67 |
| 30 | 98.372 | 97.1 | 97.04 |
| 40 | 98.706 | 97.23 | 97.19 |
| 50 | 99.002 | 97.31 | 97.19 |
| 60 | 99.196 | 97.38 | 97.16 |
| 70 | 99.34 | 97.36 | 97.26 |
| 80 | 99.438 | 97.39 | 97.33 |
| 90 | 99.542 | 97.36 | 97.34 |
| 100 | 99.612 | 97.41 | 97.4 |

Conclusion

It can be observed from the results that the multi class regression method of logistic regression yields overall higher accuracy of prediction than binary class regression.

In Support Vector Machine, a Radial Basis Function kernel yields better prediction percentage values than a Linear kernel. The accuracy increases as the value of parameter C increases.