Project Assignment 3
Project Report

First Name: Viswa Nihar
Last Name: Nukala
UBID: 50414392

First Name: Nithin Sastry
Last Name: Tellapuri
UBID: 50365532

First Name: Amartya
Last Name: Banerjee
UBID: 50425019

Introduction:

In this assignment, we need to implement logistic regression and SVM and compare the results using the MNIST dataset.

Problem 1:

Binary Logistic Regression:

Below are the results after running Binary Logistic Regression on training data, validation data and test data:

	Accuracy	Error
Training	92.75%	7.25%
Validation	91.44%	8.56%
Test	92.0%	8%

We can see that the training error is less than test data, which means that it doesn't perform as well on unknown data as it performs on known data, but the error difference is too low and is expected in this case is expected as the pattern of the data is same for all the sets.

Following is the error for train_data for each class:

Class	Error value
0	0.020817706711114594
1	0.021744985928195102
2	0.061687169921200276
3	0.07532286800889788
4	0.04433354722286809
5	0.08360044152638227
6	0.03393826581777925
7	0.04329863789915351
8	0.1100124519008449
9	0.0965520275514656

Following is the error for test_data for each class:

Class	Error value
0	6.596245535331479
1	67.1768541574032
2	217.28579131178395
3	282.53198772568857
4	305.1133386042808
5	452.0703330496352
6	578.1217386109581
7	650.3031249011111
8	725.6133921477037

9	760.2991383566518

Multi-class Logistic Regression:

Below are the results after running Multi-class Logistic Regression on training data, validation data and test data:

	Accuracy	Error
Training	93.03200000000001%	6.97%
Validation	92.47%	7.53%
Test	92.47999999999999%	7.53%

As we have seen in the BLR, in a similar way the testing error is less when compared to the test data and hence it performs well on the know data rather than the unknown but the difference in error is very low and is expected as the pattern of the data is the same for all the sets.

We observe that Multi-class Logistic regression works better than binary Logistic Regression. The test data error is lesser compared to the prior method.

Support Vector Machines:

Using linear kernel (with all other parameters kept as default):

Following is the accuracy for training, validation, and test data:

	Accuracy
Training	92.69800000000001%
Validation	91.25%
Test	92.47999999999999%

We can see that linear kernel works almost as the linear regression i.e., Binary Logistic Regression.

Using Radial Basis function with ɣ = 1 (with all other parameters kept as default):

Following is the accuracy for training, validation, and test data:

	Accuracy
Training	100.0%
Validation	14.99%
Test	16.61%

This overfits the model as the training error is 0% and the test accuracy is 16.61% which is too less. This is due to the high value of ɣ.

Using radial basis function with value of gamma setting to default (all other parameters are kept default):

Following is the accuracy for training, validation, and test data:

	Accuracy
Training	92.18%
Validation	92.1%
Test	92.61%

This setting performs better than the model set with $\gamma = 1$.

Using radial basis function with value of gamma setting to default and varying value of C (1, 10, 20, 30, · · ·, 100):

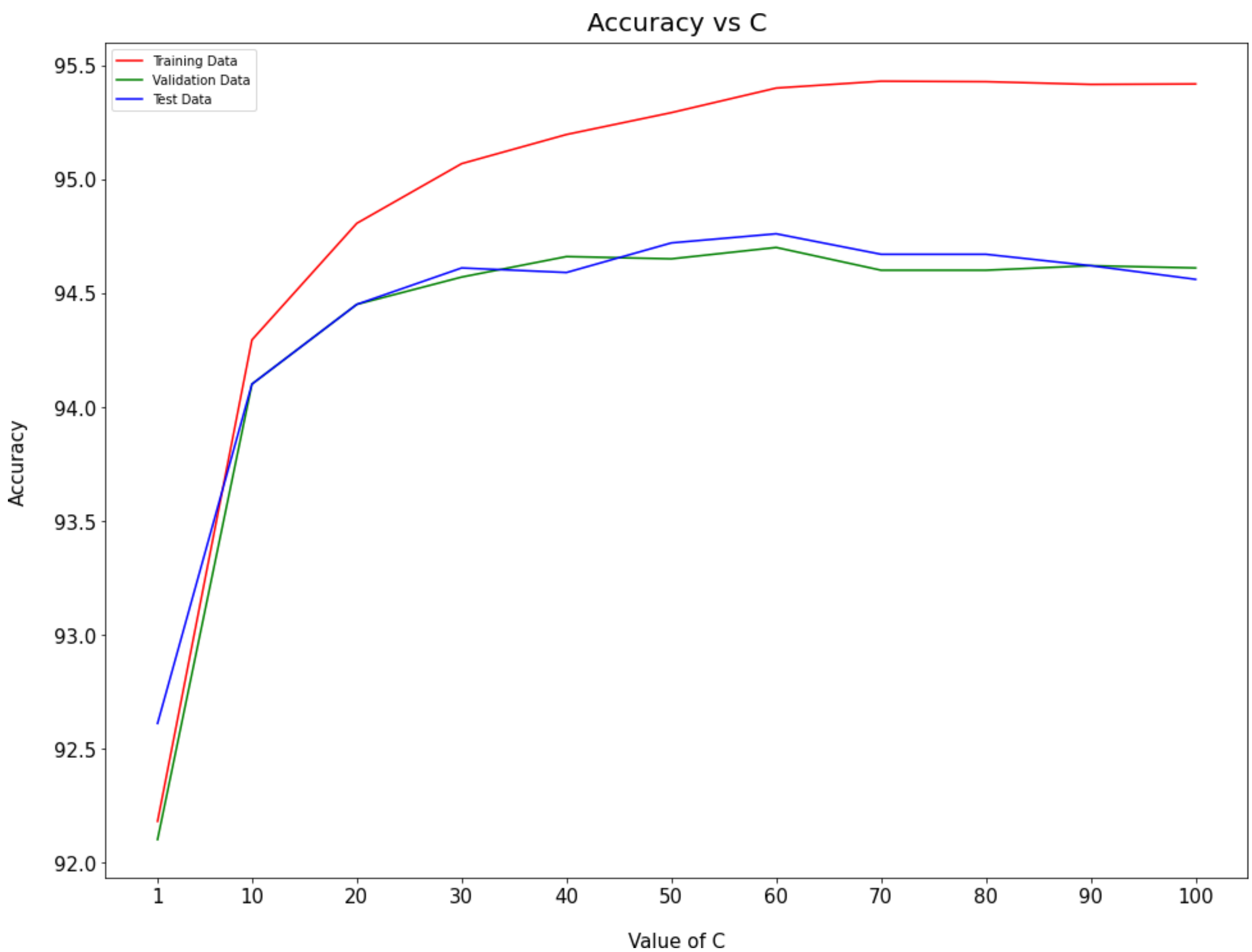
Following are the accuracies for training, validation, and test data for different values of C for partial data set:

C	Training	Validation	Test
1	92.18	92.1	92.61
10	94.294	94.1	94.1
20	94.806	94.45	94.45
30	95.068	94.57	94.61
40	95.196	94.66	94.59
50	95.292	94.65	94.72
60	95.4	94.7	94.76
70	95.43	94.6	94.67
80	95.428	94.6	94.67

90	95.416	94.62	94.62
100	95.418	94.61	94.56

For C=60, test error is the least and the accuracy is highest.

Following is the plot for C vs accuracy value for training, validation, and test data set:



Hence, we can say that the optimal C is 60 and optimal accuracy is 94.76% when keeping the Ɣ default.

Following are the accuracies for the whole data set with optimal C and gamma default setting:

	Accuracy
Training	99.196%
Validation	97.38%
Test	97.16%

So, we can say that the data fits better in when we use a non-linear model.