University at Buffalo
School of Engineering and Applied Sciences

CSE 574 – Intro to Machine Learning
Fall 2022

Assignment #3

Prepared by

Antarpreet Kaur - 50486342
Kamala Krishnan - 50478764
Sakshamdeep Singh - 50475857

December 12, 2022

TABLE OF CONTENTS

1) Binary Logistic Regression

Following are the results after running Binary Logistic Regression on training, validation, and test data.

SET	ACCURACY	ERROR
Training	92.696%	7.304%
Validation	91.479%	8.521%
Testing	92.01%	7.99 %

2) Multi – Class Logistic Regression

Following are the results after running Multi-class Logistic Regression on training, validation, and test Data.

SET	ACCURACY	ERROR
Training	93.094%	6.906%
Validation	92.47%	7.53%
Testing	92.52%	7.48%

3) Multi-class strategy (MLR) with one-vs-all (BLR) strategy comparisons

SET	BLR ACCURACY	MLR ACCURACY
Training	92.696%	93.094%
Validation	91.479%	92.47%
Testing	92.01%	92.52%

- In multiclass logistic regression we classify all the 10 classes of MNIST dataset at once, where in one-vs-all (BLR) we only classify one class with respect to all other at a time, so multiclass has less time complexity.
- Multiclass Logistic Regression performs better than the Binary Logistic Regression because the parameters are estimated independently.

4) Support Vector Machine (SVM)

I. Using a linear kernel

SET	ACCURACY
Training	99.84%
Validation	91.11%
Testing	92.04%

II. Using radial basis function with value of gamma setting to 1

SET	ACCURACY
Training	100.0%
Validation	10.00%
Testing	11.35%

Gamma is a kernel coefficient. The bigger the gamma is, the more 'linear' the decision boundary will be. Using gamma value of 1 makes the model overfit on the training data which gives strikingly poor results on unseen data, i.e., validation and test data.

III. Using radial basis function with value of gamma setting to default

SET	ACCURACY
Training	98.73%
Validation	96.0%
Testing	96.18%

Gamma is a kernel coefficient. The default value of gamma means that it will use 1/(n_features * X.var()) as value of gamma. This model performs really well on the training as well as the unseen data.

IV. Using radial basis function with value of gamma set to default and varying value of C (1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100)

C is a regularization parameter, and the strength of the regularization is inversely proportional to C. We iterate through the C values and record the value that performs best on the unseen validation data. We then use the optimal C and test on the whole dataset.

Below are the results for different values of C on training, validation, and testing data.

C	TRAINING ACCURACY	VALIDATION ACCURACY	TESTING ACCURACY
1	98.79%	95.94%	96.33%
10	100%	96.50%	96.73%
20	100%	96.49%	96.76%
30	100%	96.49%	96.76%

40	**100%**	**96.49%**	**96.76%**
50	**100%**	**96.49%**	**96.76%**
60	**100%**	**96.49%**	**96.76%**
70	**100%**	**96.49%**	**96.76%**
80	**100%**	**96.49%**	**96.76%**
90	**100%**	**96.49%**	**96.76%**
100	**100%**	**96.49%**	**96.76%**

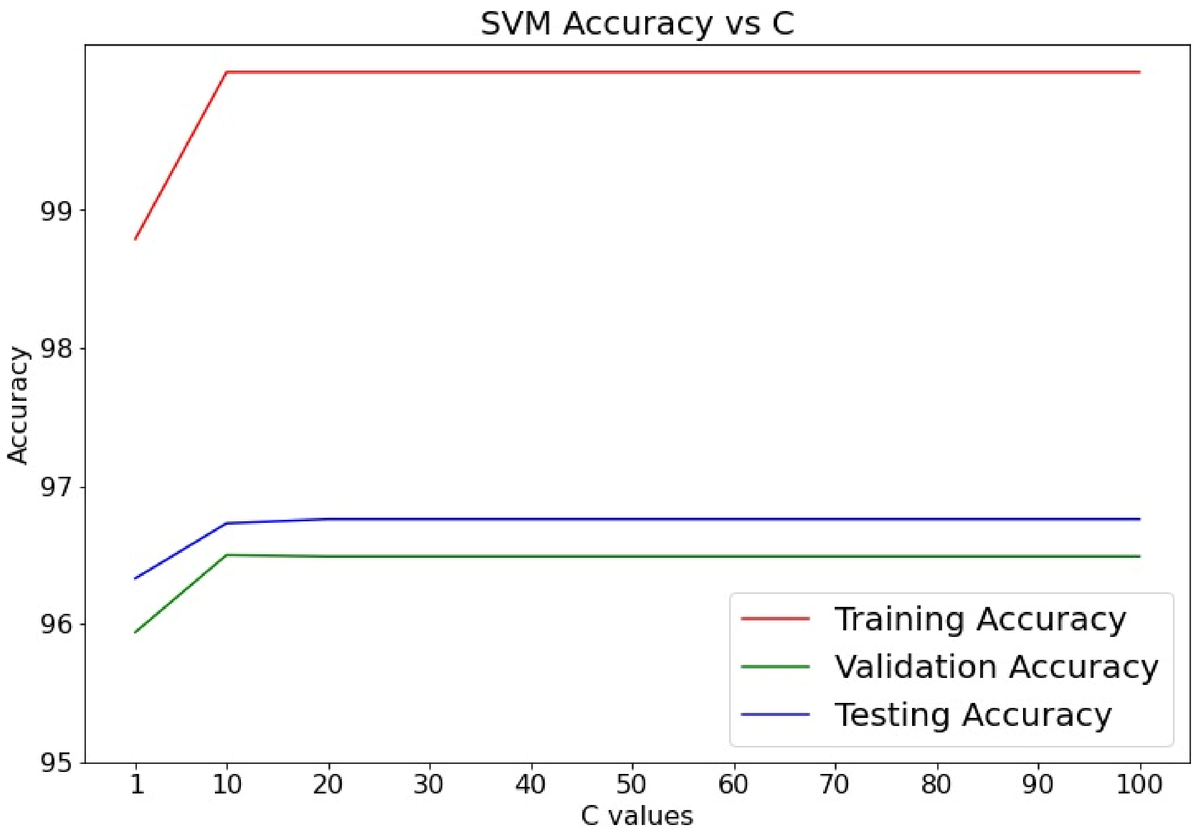
C = 10 gives us the best result on the validation data set.

RBF with FULL Training set with the optimal C :

KERNAL	Optimal C	Training Accuracy	Validation Accuracy	Testing Accuracy
RBF	**10**	**99.97%**	**98.45%**	**98.34%**

V. Accuracy Graph

Plot for accuracy vs the C values. We know that C is a regularization parameter with the strength of the regularization inversely proportional to the value of C. From the graph we can see that for C>10, the training accuracy is close to 100% which implies that the model overfit the training data and does not further improve its performance on the unseen data. We'll iterate C in the range of 1 -> 9 to further check for the value that optimizes on the unseen data.



VI. Using radial basis function with value of gamma setting to default and varying value of C (1, 2, 3, 4, 5, 6, 7, 8, 9)

To get a better understanding, we iterate through the C values from 1 → 9 and record the optimum setting and plot the accuracy obtained on each of training, validation, and testing dataset with respect to various values of C:

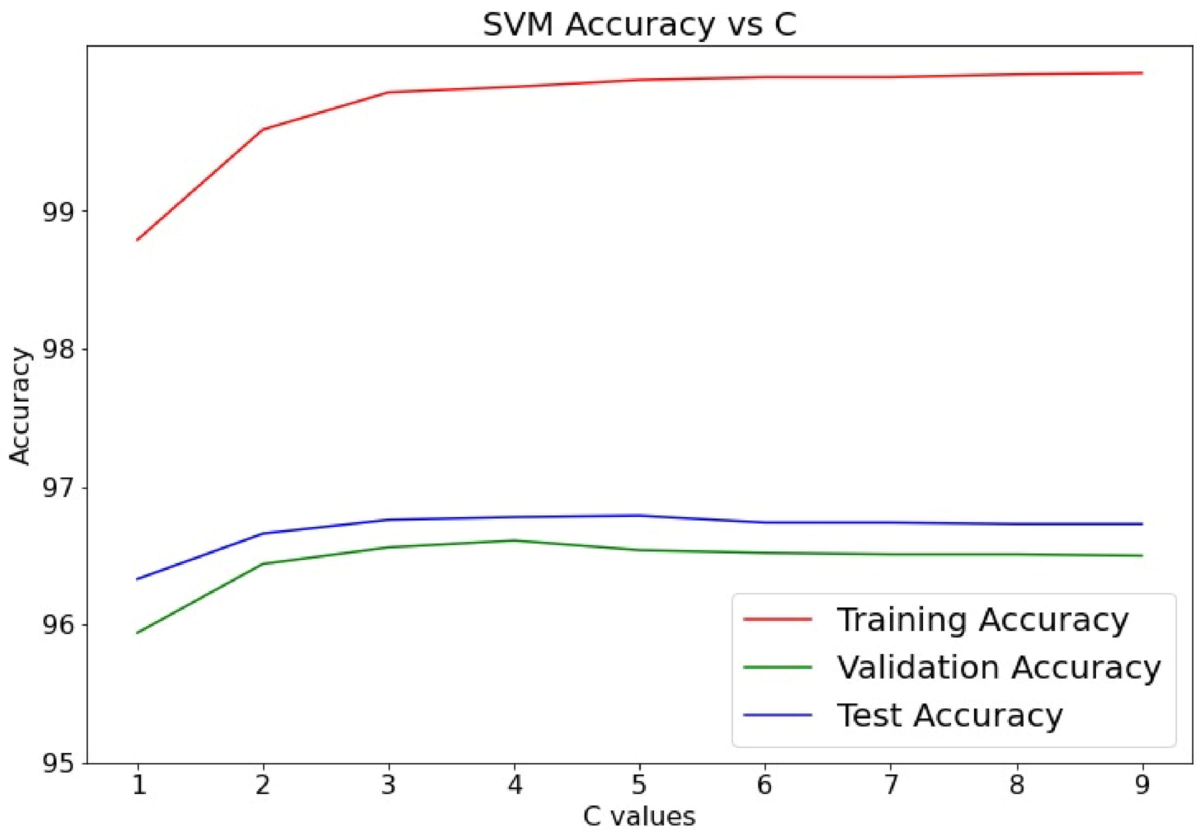
Below are the results for different values of C on training, validation, and test data.

C VALUE	TRAINING ACCURACY	VALIDATION ACCURACY	TESTING ACCURACY
1	98.79%	95.94%	96.33%
2	99.59%	96.44%	96.66%
3	99.86%	96.56%	96.76%
4	99.9%	96.61%	96.78%
5	99.95%	96.54%	96.79%
6	99.97%	96.52%	96.74%
7	99.97%	96.51%	96.74%
8	99.99%	96.51%	96.73%
9	100%	96.5%	96.73%

C = 4 gives us the best result on the validation data set.

VII. Accuracy Graph

Plot of accuracy obtained on each of Training, Testing and Validation dataset with respect to various values of C:



5) Observations

a. SVM performs better than the Logistic Regression. It is due to the fact that the logistic regression considers all the points withing a dataset for finding a decision boundary whereas SVM just works with few data points namely the support vectors and maximizes the distance between them.

b. As seen from the accuracy graph, as C increases, the regularization strength decreases, and the training accuracy increases implying overfitting.

c. In the SVM, we have a trade-off between complexity of decision rule and frequency of error by changing the parameter C.

d. SVM with gaussian kernel performs better than the linear kernel because our dataset is not linearly separable.