CSE 574 Fall 2021 Introduction to Machine Learning

Programming Assignment 3

Classification and Regression

Group 72

Ashutosh Shailesh Bhawsar, 50416025

Jayesh Kishor Suryavanshi, 50417114

Ankit Kumar, 50412887

1: Introduction

This assignment is about implementing Logistic Regression (using one-vs-all strategy) to classify handwritten digit images into corresponding labels accurately using one versus all strategy. Here, we built one binary classifier for each class i.e., total 10 classifiers to distinguish a given class from all other classes. We also developed multiclass logistic regression classifier and support vector machine for handwritten digits classification on the famous MNIST Dataset.

2: Implementation of Binary Logistic Regression (BLR)

Set	Accuracy	Error
Training	92.75%	7.25%
Validation	91.36%	8.64%
Testing	92.04%	7.96%

These are the results after running Binary Logistic Regression on Training, Validation and Testing Datasets. We can see that the Training error is less than Testing error. Hence, the conclusion is that this Linear model performs better on the seen data, but when it gets the unseen dataset, it gives a little more error. However, the difference is negligible. That is natural for any Linear Model and there is not huge difference in error as the data is of similar type in all sets.

3: Implementation of Multi-class Logistic Regression (MLR)

Set	Accuracy	Error
Training	93.27%	6.73%
Validation	92.55%	7.45%
Testing	92.54%	7.46%

These are the results after running Multi-class Logistic Regression on Training, Validation and Testing Datasets. We can see that the Training error is slightly less than Testing error. So, we can infer that this Linear model performs better on the seen data, but when it gets the unseen dataset it gives a little more error. This is natural for any Linear Model and the errors are almost equal as the pattern of data is same for all the sets.

4: Performance difference between BLR and MLR

Set	BLR Accuracy	MLR Accuracy
Training	92.75%	93.27%
Validation	91.36%	92.55%
Testing	92.04%	92.54%

The observations are that the accuracy of the multi-class classification is better than the one-vs-all classification. The reason being the parameters are estimated independently in multi-class which helps to prevent wrong classification, hence results in better classification. Furthermore, in multi-class logistic regression we classify all the classes (total 10) of MNIST dataset at once, whereas in binary logistic regression, we only classify one class with respect to all other at a time, so multi-class has lesser chances of overlapping as well as lesser time complexity.

5: Support Vector Machines

➢ 5.1: Linear Kernel

Set	Accuracy
Training	92.48%
Validation	90.90%
Testing	91.49%

We can conclude from the results that Linear Kernel works like a linear model, as the results are almost same as the previous linear model we trained.

➢ 5.2: Radial Basis Function

Using value of Gamma as 1, we get poor results on test data as this high value of gamma helps in overfitting the training data and hence, we get 100% Training accuracy as shown in table below:

Set	Accuracy
Training	100.00%
Validation	15.95%
Testing	17.90%

Next, we use Radial Basis Function with value of gamma setting to default and all other parameters set to default and observed following results:

Set	Accuracy
Training	92.02%
Validation	91.95%
Testing	92.36%

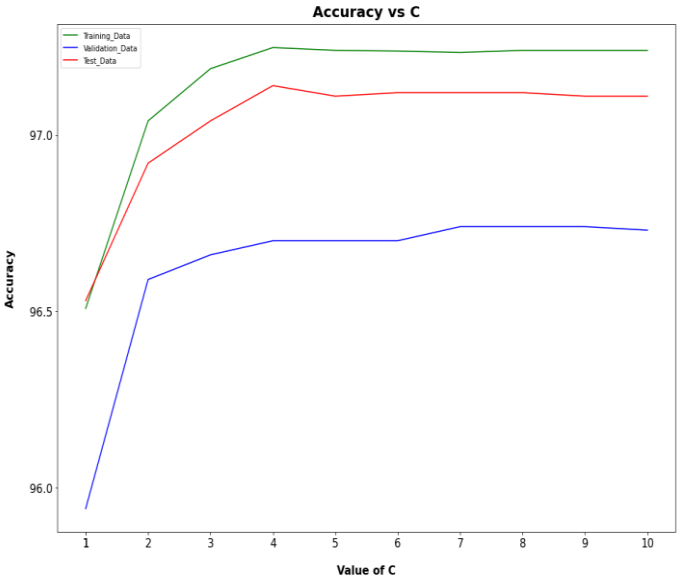
Now we iterate through the C values and observe the optimal setting of variables and then test the whole data on the current settings. The C variable controls the importance that we are giving to the Slack variable. Hence, we observe a trade-off between the width of the margin and C value. Below are the results through the varying values of C between 1 to 10 with increment of 1.

C	Training	Validation	Testing
1	96.50%	95.94%	96.53%
2	97.04%	96.59%	96.92%
3	97.18%	96.66%	97.04%
4	97.24%	96.70%	97.14%
5	97.24%	96.70%	97.11%
6	97.23%	96.70%	97.11%
7	97.24%	96.74%	97.12%
8	97.24%	96.74%	97.12%
9	97.24%	96.74%	97.11%
10	97.24%	96.73%	97.11%

Using the optimal parameters, the results observed for whole dataset are (Gamma=default):

Kernel	C	Training Accuracy	Validation Accuracy	Testing Accuracy
RBF	7	99.34%	97.36%	97.26%

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



The observations are that non-linear model gives better results, and our dataset is non-linear as well.