CSE4/573 FALL 2022 Introduction to Machine Learning

Programming Assignment 3

Classification and Regression

Team members:

- Ayush Rawat – *50478823*
- Sai Sharan Thirunagari – *50483836*
- Nikhil Naveen Chandra – *50469151*

Binary Logistic Regression:

SET	ACCURACY	ERROR
Training	92.686	7.314
Validation	91.46	8.54
Testing	91.95	8.05

From the above we observe that there is a minimal decrease in the accuracy by 0.736% or minimal increase in by 0.736% with testing data when compared with training data. Which is generally the case with any linear ML model.

Multi-Class Logistic Regression:

SET	ACCURACY	ERROR
Training	93.274	6.726
Validation	92.55	7.45
Testing	92.54	7.46

Similar to Binary LR it is observed that there is minimal decrease in the accuracy of 0.734% or minimal increase in the error with testing data when compared with training data. Which is the general case with any linear ML model

Comparison of Accuracy between BLR and MLR:

SET	BLR ACCURACY	MLR ACCURACY
Training	92.686	93.274
Validation	91.46	92.55
Testing	91.95	92.54

From the above table after comparison of accuracy between BLR and MLR in terms of Training, Validation and Testing we can conclude that overall MLR accuracy is better. It is because the MLR classify all the classes at once whereas BLR considers one vs all approach to classify which increases time complexity and chance for overlapping. One another reason is that the parameters are estimated independently which avoids wrong classification

Support Vector Machine:

i) When we use Linear Kernel:

SET	ACCURACY
Training	92.586
Validation	91.54
Testing	91.79

The results are almost same as BLR and MLR when linear kernel is used.

ii) When we use Radial Basis Function:

i) (Gamma=1):

SET	ACCURACY
Training	27.52
Validation	10.00
Testing	11.35

The use of radial basis function with the gamma as 1 result in very poor accuracy in all three phases with training data being high compared to validation and testing.

ii) (Gamma=default):

SET	ACCURACY
Training	92.08
Validation	91.97
Testing	92.5

With the gamma as default, we observe that it is approximately same as linear kernel as well as MLR and BLR.

iii) Changing C= 10,20,30…100:

C	Training	Validation	Testing
1	96.55	96.16	96.22
10	97.238	96.57	96.62
20	97.23	96.54	96.61
30	97.23	96.53	96.61
40	97.23	96.53	96.61
50	97.23	96.54	96.61
60	97.23	96.54	96.61
70	97.23	96.54	96.61
80	97.23	96.54	96.61
90	97.23	96.54	96.61
100	97.23	96.54	96.61

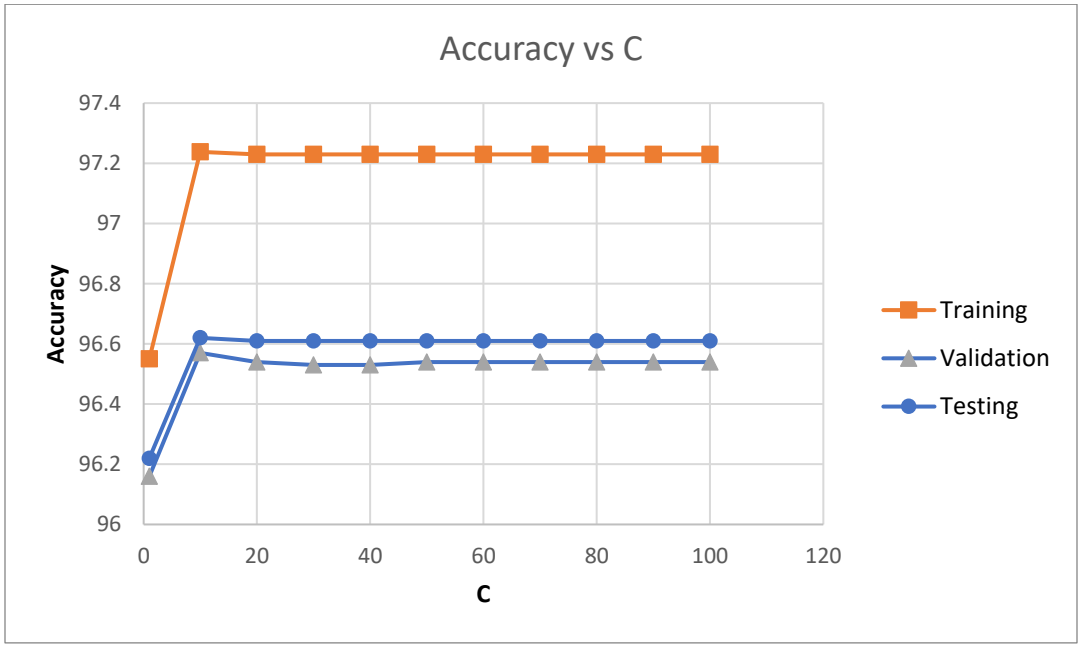
It can be observed that after training 10000 images with different C values the accuracy stagnated after C=40. So, C = 50 can be considered as best value with gamma being default.

Result of training whole dataset:

Kernel	C	Training	Validation	Testing
Default	50	99.307	99.256	99.429

The optimal parameter results in accuracy of nearly 100% after training with entire data.

Plot of Accuracy vs C in terms of testing, validation and testing:



Conclusion:

It can be concluded that the dataset is non-linear as the results are better with non-linear model.