CSE 574 Introduction to Machine Learning

Programming Assignment 3

Classification and Regression

<u>Group 11:</u>

Ahut Gupta (50169273)

Rachna Shivangi (50169516)

Vandana Chokkam (50167945)

1. Logistic Regression

Logistic Regression is a statistical method for classifying a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured as a probability. Here, we have used a one-vs-all classifier and built 10 binary classifiers for each of the output class. Finally, for predicting a class, we selected the class for which the probability was highest as the output class for that example.

Experimental Results using 10-binary classifiers:

Data Set	Accuracy (%)
Training	86.186
Validation	85.3
Testing	85.42

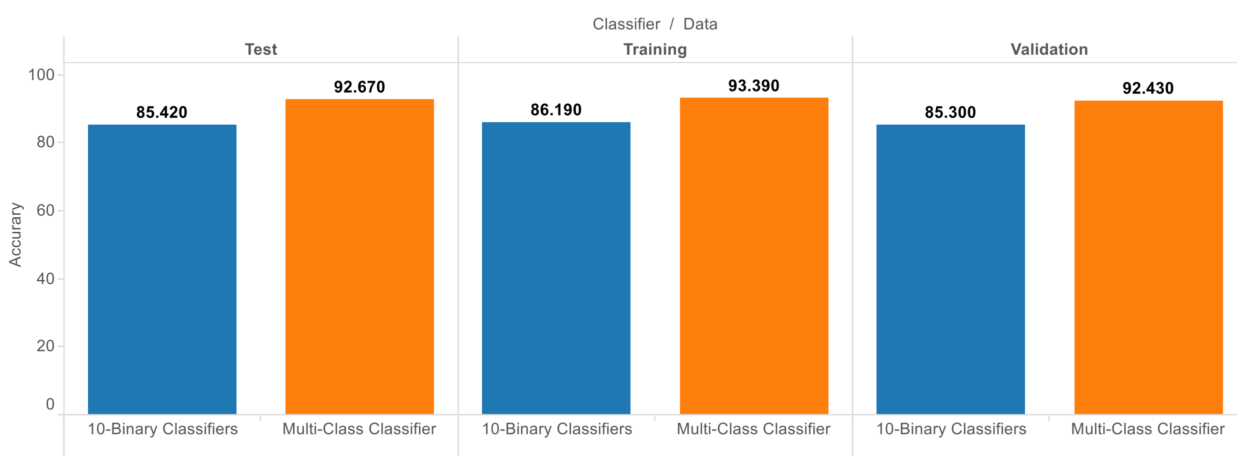
2. Multi-class Logistic Regression

The multi-class logistic regression uses an exponential function (non-linear) to preprocess the scores and then compute the percentage of each score in the sum of all the scores which can be interpreted as the probability of each class for one sample.

Experimental Results using multi-class logistic classifier:

Data Set	Accuracy (%)
Training	93.39
Validation	92.43
Testing	92.67

Comparing the two methods of Logistic Regression



Multi-class (Softmax) logistic regression works better for mutually exclusive classes as is the case with MNIST digit classification. Accuracy obtained is higher for all datasets using the multi-class classifier.

3. Support Vector Machine

A Support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification. A good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (margin), since in general the larger the margin the lower the generalization error of the classifier.

We used the SVM package by SKLearns in our project, with various parameters.

1. Experimental Results using Linear Kernel (all other parameters – default)

Here we use a Linear Kernel method to classify our data. This method works well if the data is linearly separable.

Data Set	Accuracy (%)
Training	97.286
Validation	93.64
Testing	93.78

2. Experimental Results using Radial Basis function (Gamma = 1; all other parameters – default)

Here we use a Radial Basis Function as our kernel method and Gamma value of 1. Gamma is the kernel coefficient. Which controls how much each feature affects the separating boundary. If γ is high, SVM tries to find highly complex surfaces to distinguish between the classes, hence overfits the training data. This is evident from our result, as we get 100% accuracy (overfitting) on the training data and very low accuracies for validation and test data.

Data Set	Accuracy (%)
Training	100
Validation	15.48
Testing	17.14

3. Experimental Results using Radial Basis function (Gamma = Default; all other parameters – default)

Here, the default value of gamma is used, which is 1/n_features. We can see that this gives better generalizability in the form of high validation and test accuracies, compared to the previous case.

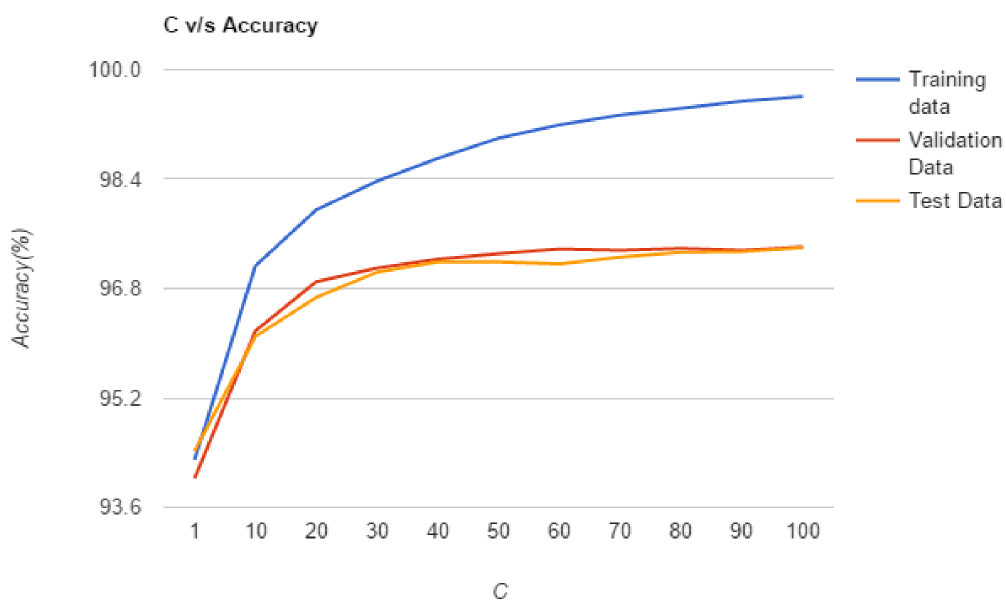
Data Set	Accuracy (%)
Training	94.294
Validation	94.02
Testing	94.42

4. Experimental Results using Radial Basis function (Gamma = Default and varying value of C)

'C' is the Penalty parameter of the error term. For low C values, we get separating planes with very large margins. Whereas for higher C values, we get low-margin separating planes.

Value of C	Training Accuracy (%)	Validation Accuracy (%)	Testing 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

Plot of Values of C (X-axis) vs. Accuracy % (Y-axis)



4. Conclusion

The SVM treats the outputs as (uncalibrated and possibly difficult to interpret) scores for each class while the Multi-Class softmax classifier gives a slightly more intuitive output (normalized class probabilities) and also has a probabilistic interpretation. The accuracies obtained from SVM(C=1) and multiclass logistic regression are comparable for the MNIST dataset.