CSE 474/574

Introduction to Machine Learning

Programming Assignment 3

Classification and Regression

GROUP 71
SIVAKUMAR PASUPATHI 50366350

PARTHIBAN RAJENDRAN 50415960

SAI PRASHANTH SELVAGANAPATHY 50419614

1)Implementation of Logistic Regression

Below is the result for the training, validation and testing set accuracy. The training set has the highest accuracy at 92.72%. The testing set has the accuracy of 91.9%. The training error is low because while training the model we are trying to modify the parameters and lower the loss function based on the training data. If the testing set is a small subset of the training data, then testing set might have lower error based on the distribution of data points.

Set	Accuracy	Error
Training Set	92.71%	7.29%
Validation Set	91.47%	8.52%
Testing Set	91.9%	8.1%

Class testing accuracy and error%:

Class	Training Accuracy%	Training Error %	Validation Accuracy%	Validation Error%	Testing Accuracy%	Testing Error%
0	97.82	2.18	97.6	2.4	98.06	1.94
1	97.94	2.06	96.9	3.1	98.23	1.77
2	91.1	8.9	88.2	11.8	89.05	10.95
3	83.67	16.33	88.5	11.5	91.28	8.72
4	93.84	6.16	93.9	6.1	93.17	6.83
5	88.26	11.74	86.6	13.4	85.43	14.57
6	96.36	3.64	95.7	4.3	94.78	5.22
7	94.18	5.82	92.2	7.8	92.5	7.5
8	87.38	12.62	84.3	15.7	87.06	12.94
9	89.2	10.8	90.4	9.6	89	11

```
Training set Accuracy:92.712%

Training set Accuracy for class:0  97.82652853950843%

Training set Accuracy for class:1  97.94496691048415%

Training set Accuracy for class:2  91.10528438886648%

Training set Accuracy for class:3  89.67062950691873%

Training set Accuracy for class:4  93.84551838083436%

Training set Accuracy for class:5  88.26057453064917%

Training set Accuracy for class:6  96.36030906872712%

Training set Accuracy for class:7  94.18803418803418%

Training set Accuracy for class:8  87.38404452690168%

Training set Accuracy for class:9  89.2099414023035%
```

```
Validation set Accuracy:91.43%

Validation set Accuracy for class:0  97.6%

Validation set Accuracy for class:1  96.89999999999999%

Validation set Accuracy for class:2  88.2%

Validation set Accuracy for class:3  88.5%

Validation set Accuracy for class:4  93.89999999999999%

Validation set Accuracy for class:5  86.6%

Validation set Accuracy for class:6  95.7%

Validation set Accuracy for class:7  92.2%

Validation set Accuracy for class:8  84.3%

Validation set Accuracy for class:9  90.4%
```

```
Testing set Accuracy:91.99000000000001%

Testing set Accuracy for class:0  98.06122448979592%

Testing set Accuracy for class:1  98.23788546255507%

Testing set Accuracy for class:2  89.05038759689923%

Testing set Accuracy for class:3  91.2871287128713%

Testing set Accuracy for class:4  93.17718940936864%

Testing set Accuracy for class:5  85.42600896860986%

Testing set Accuracy for class:6  94.78079331941545%

Testing set Accuracy for class:7  92.50972762645915%

Testing set Accuracy for class:8  87.06365503080082%

Testing set Accuracy for class:9  88.9990089197225%
```

2) Multi-class Logistic Regression

The result of Training, validation and testing set accuracy are as given below. The testing set accuracy is 92.48%. The One vs All method requires a lot of computation because for each class we need to create a separate model. The One vs All method also requires a lot of data to train.

Set	Accuracy	Error
Training Set	93.282%	6.718%
Validation Set	93.42%	6.58%
Testing Set	92.48%	7.52%

3)SVM

SVM With Linear kernel

The accuracy of SVM with linear kernel is almost equal to that of the linear model and as given below. The testing data accuracy is at 91.5%

Set	Accuracy	Error
SVM training	99.81%	0.19%
Training Set	92.59%	7.41%
Validation Set	91.6%	8.4%
Testing Set	91.5%	8.5%

SVM RBF

SVM RBF gives very low accuracy of 18.73% for the testing data.We can see that SVM training accuracy is at 100% due to overfitting of the training data set.

Set	Accuracy	Error
SVM training	100%	0%
Training Set	33.614%	66.386%
Validation Set	16.89%	83.11%
Testing Set	18.73%	81.27%

SVM RBF with Default Gamma

Set	Accuracy	Error
SVM training	92.85%	7.15%
Training Set	91.86	8.14%
Validation Set	92.08	7.92%
Testing Set	92.25%	7.75%

```
----------------------------------------SVM with Linear Kernel----------------------------------------

SVM Training Data Accuracy : 99.81%

Training Data Accuracy : 92.596%

Validation Data Accuracy : 91.60000000000001%

Testing Data Accuracy : 91.5%
----------------------------------------SVM RBF ; Gamma = 1----------------------------------------

SVM Training Data Accuracy : 100.0%

Training Accuracy : 33.614%

Validation Accuracy : 16.89%

Testing Accuracy : 18.73%
----------------------------------------SVM RBF ; Gamma = default(0.0)----------------------------------------

SVM Training Data Accuracy : 92.85%

Training Accuracy : 91.86%

Validation Accuracy : 92.08%

Testing Accuracy : 92.25%
```

SVM RBF with Default Gamma and varying C

Below is the accuracy result for SVM, Training, Validation and Testing Respectively for varying value of C. The C parameter is used to determine how much misclassification we need to avoid while training.

The margin becomes smaller as we increase C so the data for SVM Training Accuracy is very high. Optimal C value is 20 with testing error 3.13%

After value of C = 10, there is no significant change in the accuracy of the model. Since the SVM training accuracy is 100 after C >= 10, the data is almost linearly separable. Therefore, increasing the C value does not always increase the testing accuracy.

C	SVM Training Accuracy%	SVM Training Error%	Training Accuracy%	Training Error%	Validation Accuracy%	Validation Error%	Testing Accuracy%	Testing Error%
1	98.61	1.39	96.422	3.578	96.09	3.91	96.26	3.74
10	100	0	97.252	2.748	96.58	3.42	96.86	3.14
20	**100**	**0**	**97.252**	**2.748**	**96.56**	**3.44**	**96.87**	**3.13**
30	100	0	97.252	2.748	96.56	3.44	96.87	3.13
40	100	0	97.252	2.748	96.56	3.44	96.87	3.13
50	100	0	97.252	2.748	96.56	3.44	96.87	3.13
60	100	0	97.252	2.748	96.56	3.44	96.87	3.13
70	100	0	97.252	2.748	96.56	3.44	96.87	3.13
80	100	0	97.252	2.748	96.56	3.44	96.87	3.13
90	100	0	97.252	2.748	96.56	3.44	96.87	3.13
100	100	0	97.252	2.748	96.56	3.44	96.87	3.13

```
-------------------------------------SVM RBF ; Gamma = default(0.0) ; C  = 1-------------------------------------------

SVM Training Data Accuracy : 98.61%

Training Accuracy : 96.422%

Validation Accuracy : 96.09%

Testing Accuracy : 96.26%
-------------------------------------SVM RBF ; Gamma = default(0.0) ; C  = 10-------------------------------------------

SVM Training Data Accuracy : 100.0%

Training Accuracy : 97.25200000000001%

Validation Accuracy : 96.58%

Testing Accuracy : 96.86%
-------------------------------------SVM RBF ; Gamma = default(0.0) ; C  = 20-------------------------------------------

SVM Training Data Accuracy : 100.0%

Training Accuracy : 97.25200000000001%

Validation Accuracy : 96.56%

Testing Accuracy : 96.87%
-------------------------------------SVM RBF ; Gamma = default(0.0) ; C  = 30-------------------------------------------

SVM Training Data Accuracy : 100.0%

Training Accuracy : 97.25200000000001%

Validation Accuracy : 96.56%

Testing Accuracy : 96.87%
```
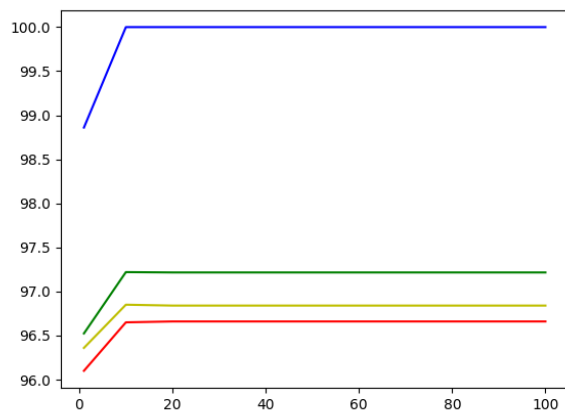
Accuracy vs C

Testing accuracy for optimal C

```
SVM Training Data Accuracy for C = 20: 100.0%

Training Accuracy for C = 20: 97.27799999999999%

Validation Accuracy for C = 20: 96.72%

Testing Accuracy for C = 20:97.06%
```