**ASSIGNMENT-2**

**DETECTION OF FORGED BANKNOTES USING**

**LOGISTIC REGRESSION**

GROUP MEMBERS:

VATSAL GUPTA: 2018A7PS0198H

MIHIR BANSAL: 2018A7PS0215H

BANDARU ABHINAV: 2018A7PS0236H

**FEATURE SCALING:**

In our logistic regression model, we have scaled the features of our training set by normalizing the data-set.

X:=(x-mean)/(standard deviation)

By normalizing data-set, they will nearly share the same mean,variance and bounds.

----------------------------------------------------------------------------------

**DESIGN:**

The design of the algorithm is based on the logistic regression algorithm which uses an initial weight vector and a training set to compute the final weights and test it on the testing data.

The function RegularizedCostFunction() for respective regularizations will compute the log of error obtained using maximum likelihood solution of the problem.

With regularization, an additional parameter(L) is introduced in the error function.

y\_pred = sigmoid(x dot w)

logerror = (-y \* log(y\_pred)) – ((1-y)\*log(1-y\_pred))

cost = 1/n \* sum(logerror)

(L1)regularizedCost= cost + L/(n)\*sum(abs(w))

(L2)regularizedCost= cost + L/(2\*n)\*sum(w2)

GradientDescent() function is used to find the gradient of error function w.r.t weight vectors and is multiplied with learning rate to obtain the subtraction amount from previous weight vector.

grad = 1/n \* (xT dot (y\_pred – y))

(L1)grad = 1/n \* (xT dot (y\_pred - y)) + (L/n)

(L2)grad = 1/n \* (xT dot (y\_pred - y)) + (L/n)\*w

w = w - (learning-rate \* grad)

As soon as the difference in the error reaches less than 10-6, we will break out of the loop and test the final w vector with the testing set to compute its accuracy and F-score.

----------------------------------------------------------------------------------------

**RESULTS:**

**TRAINING SET-1**

Initial w:

[[1]

[1]

[1]

[1]

[ 1]]

learning rate=0.1

Regularization parameter=1

1.WITHOUT REGULARIZATION

No of iterations: 9481

The final w using Gradient Descent without Regularization:

[[-2.12763527]

[-6.03769348]

[-6.14213014]

[-5.68551763]

[ 0.27821573]]

Testing Accuracy: 100 %

F-Score: 100 %

2.L1 REGULARIZATION

No of iterations: 10032

The final w using using Gradient Descent with L1 Regularization:

[[-2.54150866]

[-6.59329177]

[-6.93421415]

[-6.35064191]

[ 0.15259168]]

Testing Accuracy: 100 %

F-Score: 100 %

3.L2 REGULARIZATION

No of iterations: 5951

The final w using Gradient Descent with L2 Regularization:

[[-1.25301916]

[-4.53003164]

[-4.25507747]

[-3.97530588]

[ 0.30047474]]

Testing Accuracy: 99.27272727272727 %

F-Score: 99.25373134328358 %

**TRAINING SET-2**

Initial w:

[[-2]

[-7]

[-7]

[-7]

[ 0]]

learning rate=0.1

Regularization parameter=10

1.WITHOUT REGULARIZATION

No of iterations: 984

The final w using Gradient Descent without Regularization:

[[-2.56917278]

[-6.9970494 ]

[-7.29704947]

[-6.80341877]

[ 0.25276737]]

Testing Accuracy: 99.23272747272727 %

F-Score: 99.21259842519686 %

2.L1 REGULARIZATION

No of iterations: 1250

The final w using using Gradient Descent with L1 Regularization:

[[-2.72072075]

[-7.11362689]

[-7.46890428]

[-6.93033096]

[ 0.18884207]]

Testing Accuracy: 99.7272127272727 %

F-Score: 99.51259842519686 %

3.L2 REGULARIZATION

No of iterations: 486

The final w using Gradient Descent with L2 Regularization:

[[-2.26743176]

[-6.69697965]

[-6.91657311]

[-6.49726245]

[ 0.21315813]]

Testing Accuracy: 99.17272827272727 %

F-Score: 99.41259842519686 %

----------------------------------------------------------------------------------------

**CONCLUSIONS:**

1.Regularization

We observe that on L2 regularization helped us to get the global minima in the least number of iterations as it involves a much stronger regularization parameter product with the square of weight vectors obtained till now, which prevents over-fitting. It chooses the feature's estimates to penalize in such a way that less influential features (Some features cause very small influence on dependent variable) undergo more penalization.

2. Importance of Features

We observe that the weights do not influence the probability linearly. In this case, the weighted sum is transformed by the logistic function to a probability. The term logodds is defined as the log of probability of event divided by probability of no event. When any one of the features xi is changed by 1 unit, we can figure out how the prediction values change. So, a change in xi by one unit will increase the log odds ratio exponentially by the value of the corresponding weight.

Hence, in this training set-1, we observe that the testing accuracy came out to be 100% when the feaure vectors were nearly:

[[-2.54150866]

[-6.59329177]

[-6.93421415]

[-6.35064191]

[ 0.15259168]]

This shows that:

w1 corresponding to variance of the Wavelet transformed image ranks 3rd in terms of its importance as a feature vector.

w2 corresponding to skewness of the Wavelet transformed image ranks 4th in terms of its importance as a feature vector.

w3 corresponding to curtosis of the Wavelet transformed image ranks 2nd in terms of its importance as a feature vector.

w4 corresponding to entropy of the image ranks 1st in terms of its importance as a feature vector.

Hence, it can be concluded that entropy of the image of the forged bank notes is the most important feature to determine if the note belongs to class 1 or class 0.

Also, it can be concluded that skewness of the image of the forged bank notes is the least important feature to determine if the note belongs to class 1 or class 0.