Machine Learning Assignment 2

Logistic Regression

Team Members→

1. Atharva Sune 2017A7PS0183H
2. Ameya Salankar 2017A7PS0182H
3. Aayush Laddha 2017A8PS0717H

Logistic Regression

Logistic regression is a binary classifier which predicts whether a class belongs to a particular class or not. In logistic regression sigmoid function is used which gives the output as a value between 0 and 1. If the value of the output is greater than 0.5 then the data point is considered to belong to the class (denoted by 1) else it does not belong to the class (denoted by 0).

Suppose a data point X has ‘n’ features (X1, X2, …..Xn), then to get the prediction we multiply these features with weights, add them and then apply the sigmoid function on it. Since there are ‘n’ features we have ‘n’ weights (W1, W2....Wn). Let the output of multiplication be called A, then

$$A = W_1X_1 + W_2X_2 + ... + W_nX_n$$

Then the output of the sigmoid function is the predicted value, let it be denoted by Z

$$Z = \sigma(A)$$

Where $\sigma(x) = \frac{1}{1 + e^{(-x)}}$

Model and Dataset

The dataset had to be scaled first before applying the model. For scaling we have subtracted the mean of a feature from all the values and then divided the result by the standard deviation of the feature. This process is called standardization.

$Z = \frac{X-\mu}{\sigma}$, where Z = new value, X = original value, μ = mean and σ = std. Deviation

Then an 80-20 split is performed on the dataset, to get the train and test split .

The weights are trained over the training data and then the model thus ready is tested on the testing data. The loss function used is **binary crossentropy,** and three different models were tested **No Regularization, L1 Regularization, L2 Regularization.**

Loss Function → $-(ylog(t) + (1-y)log(1-t))$

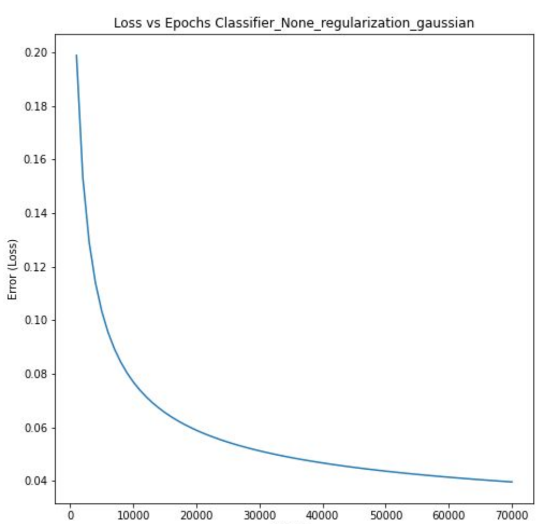
L1 Regularization → $-(ylog(t) + (1-y)log(1-t)) + \lambda(||w||)$

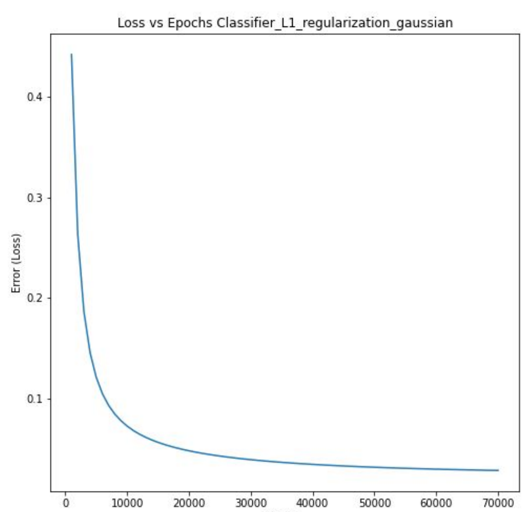
L2 Regularization → $-(ylog(t) + (1-y)log(1-t)) + \frac{1}{2}\lambda(||w||^2)$

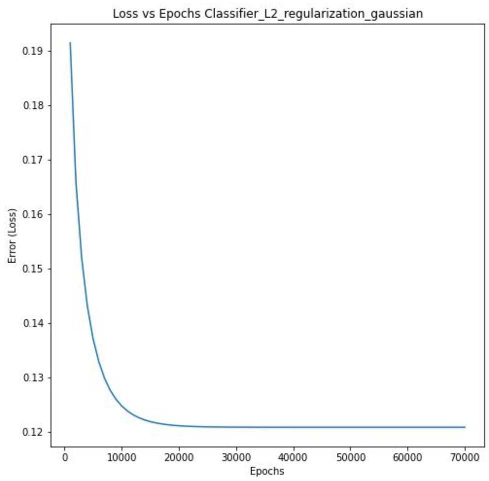
Different weight initializations were tried **Zero, Normal, Uniform.**

Gaussian Distribution plots

The following represent the graphs for all three loss functions in order (none, L1, L2)

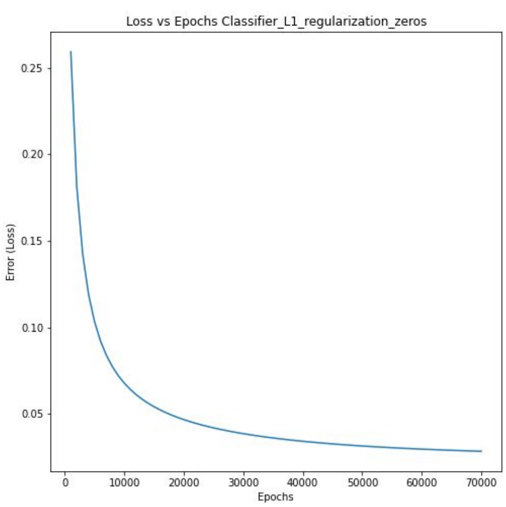


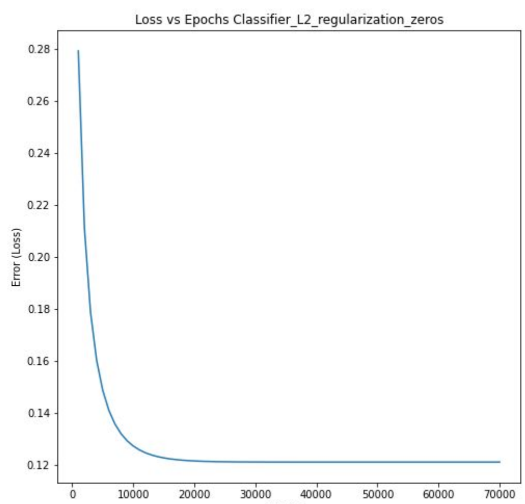


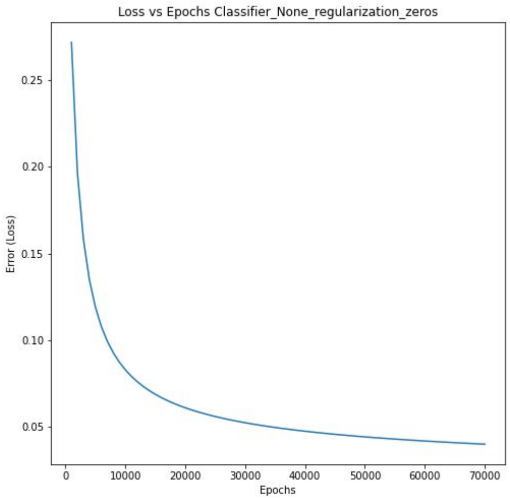


Zero initialization plots

The following plots represent the plots for zero initialization of weights for all the loss functions in order (None, L1, L2)

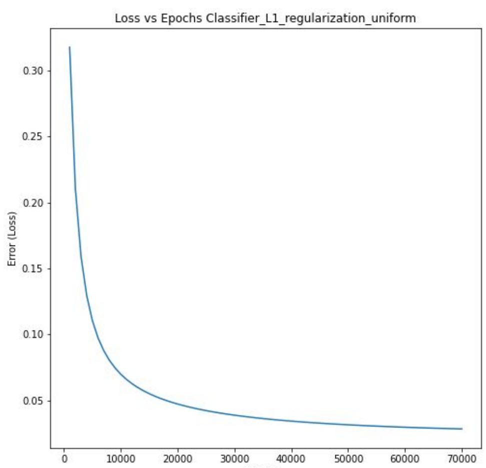


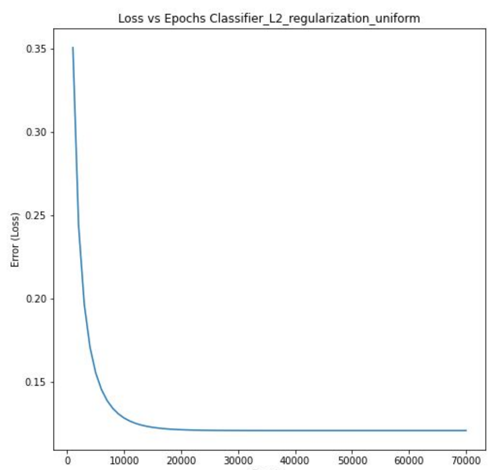


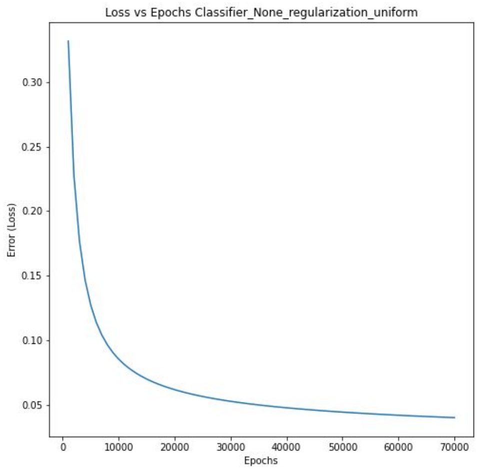


Uniform Initialization Plots

The following represent the plots for uniform initialization of weights (between 0 and 1) for all the three loss functions







Accuracy F-Score Table

	Accuracy (Train, Test)	F-score (Train, test)
None regularization Gaussian	(98.72, 99.27)	(97.9, 99.6)
L1 regularization Gaussian	(98.8, 99.2)	(98.0, 99.6)
L2 regularization Gaussian	(97.71, 97.10)	(96.27, 98.53)
None regularization Zeros	(98.7, 99.2)	(97.9, 99.6)
L1 regularization Zeros	(98.8, 99.2)	(98.0, 99.6)
L2 regularization Zeros	(97.7, 97.10)	(96.2, 98.529)
None regularization Uniform	(98.7, 99.2)	(97.9, 99.6)
L1 regularization Uniform	(98.8, 99.2)	(98.07, 99.6)
L2 regularization Uniform	(97.7, 97.10)	(96.2, 98.529)

From observing the weights it can be said that the first 3 features are the dominant ones in deciding the class of the sample, as they have a greater value whereas the 4th weight always has a value close to zero.