**Machine Learning Assignment 2**

**Logistic Regression**

Logistic regression is a classification algorithm where we divide the dataset into two or more sections so as to classify them. For the given assignment, we have to classify bank notes depending on whether they’re forgeries or not.

We are also required to incorporate L1 regression as well as L2 regression into our model. In addition we will report the F Score as well as the accuracy of the model as well.

The data was scaled using a standard normal function so as to make the distribution aligned with a standard normal curve. The impact of scaling on the accuracy and the F Score was mininal.

Among the four features given in the dataset the accuracy and the F Score as follow(for No regularization data):

|  |  |  |
| --- | --- | --- |
| **Feature** | **F Score** | **Accuracy** |
| **Variance** | 0.8765 | 87.86% |
| **Skewness** | 0.7877 | 76.13% |
| **Curtosis** | 0.8570 | 84.52% |
| **Entropy** | 0.9191 | 93.79% |

Hence, from the above analysis it can be concluded that the Entropy of the image is the most important feature of a model as it gives the highest accuracy and F Score value amont all the four given feature values.

Now we test for different learning rates and weight initializations in the model.

|  |  |  |
| --- | --- | --- |
| Weight Initializations | F Score | Accuracy |
| Uniform | 0.9133 | 93.05% |
| Gaussian | 0.8372 | 83.49% |

Hence, the uniform weight intialization gives us a better estimate than the gaussian analysis.

Now for the learning rates, we have

|  |  |  |
| --- | --- | --- |
| Learning Rate | F Score | Accuracy |
| 0.001 | 0.91283 | 93.24% |
| 0.01 | 0.91087 | 92.79% |
| 0.1 | 0.92498 | 93.299 |

Thus it can be inferred that the learning rate doesn’t have any discernible impact on the accuracy and the F score value of the logistic regression, irrespective of whether it is large or small.

The chosen feature for classification is the entropy value, along with uniform weight intialization as well as a learning rate of 0.01.

|  |  |  |
| --- | --- | --- |
|  | **F\_Score** | **Accuracy** |
| **No Regularization** | 0.907588 | 93.53% |
| **L1 Regularization** | 0.909481 | 93.18% |
| **L2 Regularization** | 0.912045 | 93.09% |

Hence, it can be seen that the accuracy of the regression decreases slightly with regularization whereas the F Value increases marginally.This can be attributed to less overfitting in the model after the implementation of the regularization exercise.

Note:- We’ve iterated 100 times with each given condition and come up with an average value for the above analysis,i.e. all the above values are the average values of a 100 iterations for the given parameters.