**Logistic Regression**

Introduction

Forging banknotes harms people and economies. Circulation of forged banknotes in the economy can lead to inflation, depreciation of real money, decrease in acceptability of paper money, individuals and companies bear losses due to non-reimbursement and increased taxes.

We built a binary classifier detector to identify whether a banknote with given attributes is forged or not. We built upon optimizing the probability of a banknote with given attributes belonging to particular class. We achieved a peak accuracy of 99.64% with our model and ordered the relative importance of features of banknotes held into account in the process.

Problem definition and Algorithm

We worked on ‘banknote authentication Data Set’ provided by UCI. Description of data as given on source - Data were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. Wavelet Transform tool were used to extract features from images.

We consider four input features for our problem of which three were variance, skewness and curtosis of Wavelet Transformed image and fourth being entropy of image. All of the features are continuous in nature. Target variable is discrete and one of class 0(false) or 1(true). Data set comprise of 1372 instances of which 762 were negative examples and 610 positives.

We implemented the binary classification algorithm – *Logistic Regression* which tries to find the best hyper-plane in 4-dimensional space of these input features which separates these 2 classes, minimizing logistic loss.

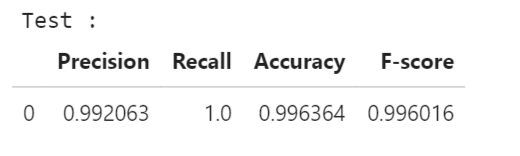
Methodology

We did 80-20 split after randomly shuffling the data set imported to a 2 dimensional *Numpy array*. This was followed by feature-scaling where we brought down the mean and variance of each feature to 0 and 1 respectively. We defined our cost function as - sum of logistic loss, L2 regularization loss (parameter λ2), and L1 regularization loss (parameter λ1 ) averaged over 1097 training examples. We used batch gradient descent optimization algorithm to find the weights [ w1 w2 w3 w4] and bias b1 minimizing cost function. We used python’s scientific computing library Numpy for ***Vectorized*** implementation allowing each epoch to pass through complete training set at once without 2 explicit for-loops which speeded up the training up to 300 times. We embedded L1 and L2 regularization in case high variance was observed. We used accuracy and f1-score as our optimizing metric and training time as satisficing metric. Hyper-parameter search was done using scaled-based grid followed by random grid.

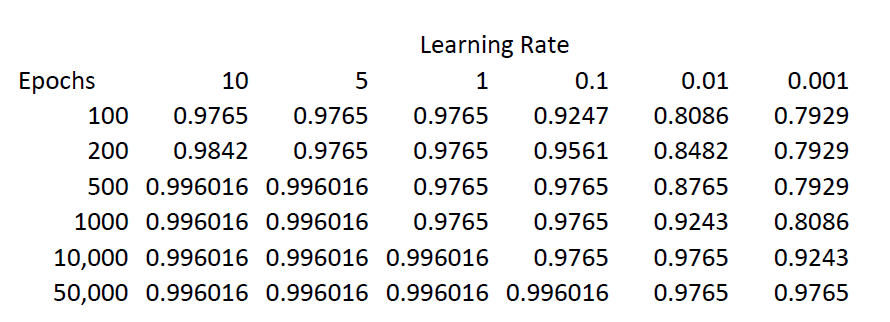
Results

1. Without regularization:

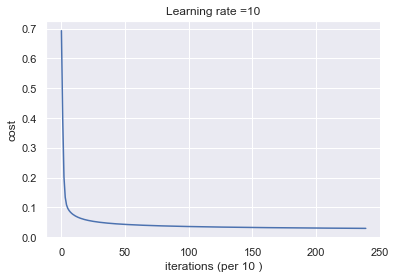
We obtained a classifier with impressive metrics as followed:



Hyper-parameter search (Metric: F1-score),



Shows that, our best hyper-parameter are:

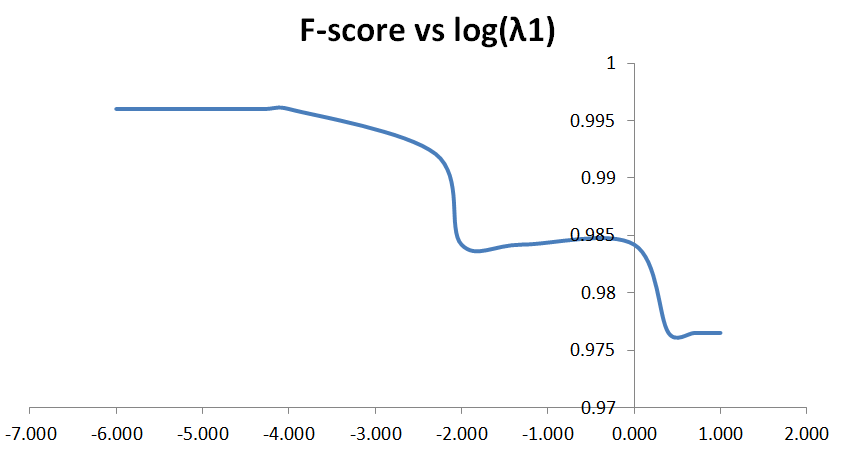


|  |  |
| --- | --- |
| Learning Rate : | 10 |
| Epochs : | 240 |

(Time taken by the model: 0.04648780822753906 sec)

*For any initialization we converged at the same point for these hyper-parameters.*

1. With L1 regularization  
     
   With above optimized hyper-parameters we plot log(λ1 ) vs F-score



1. With L2 regularization

With above optimized hyper-parameters we plot log(λ1 ) vs F-score

Feature importance:

In order to make deduction about the relative importance of the features in model for prediction we removed the standardization and regularization. We again got the same metrics of the model with same hyperparameters but different weights. The weights of the model are:

|  |  |
| --- | --- |
| **Feature** | **Weight** |
| Variance of Wavelet Transformed Image | -12.53632558 |
| Skewness of Wavelet Transformed Image | -7.02270101 |
| Curtosis of Wavelet Transformed Image | -8.32984795 |
| Entropy of Image | -1.3072809 |

Larger negative value of weights signify higher importance in the prediction of negative class. This can be seen from definition of logisitic loss.

Therefore, Variance of Wavelet Transformed Image is the most importance feature in our model.

Different weight initialization made no differenence to permformance of the model in any metric since we are using such high learning rate and our loss function is convex and bowl shaped we always land at the same optimal point.

The classifier surpasses human level performance.

There is very low bias in the model since it is very close to Bayes optimal error.

Regularization hurt the performance of the model since there is no variance problem and the model generalizes well from training set to development/test set.

References

1. <https://towardsdatascience.com/only-numpy-implementing-different-combination-of-l1-norm-l2-norm-l1-regularization-and-14b01a9773b>
2. <https://towardsdatascience.com/only-numpy-implementing-different-combination-of-l1-norm-l2-norm-l1-regularization-and-14b01a9773b>
3. [https://web.archive.org/web/20070616172012/http://wfhummel.cnchost.com/counterfeiting.html](https://web.archive.org/web/20070616172012/http:/wfhummel.cnchost.com/counterfeiting.html)
4. <https://towardsdatascience.com/model-based-feature-importance-d4f6fb2ad403>
5. <https://medium.com/ml-ai-study-group/vectorized-implementation-of-cost-functions-and-gradient-vectors-linear-regression-and-logistic-31c17bca9181>