ML Assignment-2

## Group Details

|  |  |
| --- | --- |
| **Tangeda Sai Sharan** | **2017A7PS0241H** |
| **Sanjiv Yelthimar Shenoy** | **2017A7PS0224H** |
| **Manish Kumar Bachhu** | **2017A7PS0036H** |

# **Logistic Regression – Question – 1**

## 1. Design Decisions:

### 1.1. Dataset Scaling:

We used “Standard Scalar” to scale the dataset, this methods scales the data so that each column has *Mean = 0, variance = 1*

We used numpy here and everywhere else to remove loops, thereby increasing the efficiency of the code

Then the dataset was randomly shuffled using numpy’s random.shuffle()

### 1.2. Train-Validate-Test Split:

The dataset was split into 3 parts

Train-Data: 70%

Validation-Data: 10%

Test-Data: 20%

### 1.3. Train and Validate:

We wrote a function called train\_and\_validate which takes a type of regularization as parameter and performs training followed by finding the best Lambda for regularization.

The code prints 3 outputs, clearly showing Normal, Ridge(L2), Lasso(L1) norms applied on a initialization distribution and accuracy is taken on test data set, those results are mentioned in the tables below

### 1.4. Determining the important feature:

After all the weights were updated, the absolute value of the weights was taken, then the maximum value of resulting absolute weights was taken as the most important feature.

Reasons for taking this assumption(Justification):

* Weights of features determine the how much the feature plays role in deciding the result
* If a weight is negative it doesn’t mean that it is not important, just that it negatively affects the results
* Hence in-order to consider this we took the absolute value of weights and then we chose the maximum value

### 1.5. Blog References supporting our approach:

* <https://machinelearningmastery.com/calculate-feature-importance-with-python/>
* <https://towardsdatascience.com/model-based-feature-importance-d4f6fb2ad403>
* <https://stackoverflow.com/questions/34052115/how-to-find-the-importance-of-the-features-for-a-logistic-regression-model>

## 2. Test Results with Different Initialization distributions:

### a)Gaussian

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Regularization | Test Accuracy | Test Fscore | Most Important Feature  (variance of Wavelet Transformed image with) | Least Important Feature  (curtosis of Wavelet Transformed image with ) |
| None | 98.545 | 0.9836 | Col No: 1 and weight: 0.3667 | Col No: 3 and weight: 0.01993 |
| Ridge | 98.545 | 0.9836 | Col No: 1 and weight: 0.06418 | Col No: 3 and weight: 0.00328 |
| Lasso | 98.181 | 0.9795 | Col No: 1 and weight: 0.0166 | Col No: 3 and weight: 0.000346 |

### b) Random:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Regularization | Test Accuracy | Test Fscore | Most Important Feature  (variance of Wavelet Transformed image with) | Least Important Feature  (curtosis of Wavelet Transformed image with ) |
| None | 99.636 | 0.9957 | Col No: 1 and weight: 0.03426 | Col No: 3 and weight: 0.00190 |
| Ridge | 99.272 | 0.9915 | Col No: 1 and weight: 0.01751 | Col No: 3 and weight: 0.001219 |
| Lasso | 99.272 | 0.9915 | Col No: 1 and weight: 0.01875 | Col No: 3 and weight: 0.000793 |

### c) Uniform:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Regularization | Test Accuracy | Test Fscore | Most Important Feature  (variance of Wavelet Transformed image with) | Least Important Feature  (curtosis of Wavelet Transformed image with ) |
| None | 99.272 | 0.9918 | Col No: 1 and weight: 0.2941 | Col No: 3 and weight: 0.01819 |
| Ridge | 99.272 | 0.9918 | Col No: 1 and weight: 0.1662 | Col No: 3 and weight: 0.01038 |
| Lasso | 99.272 | 0.9877 | Col No: 1 and weight: 0.04238 | Col No: 3 and weight: 0.00278 |

### Finding the Best Lambda for Regularization:

|  |  |  |
| --- | --- | --- |
| Weight Initialisation | Ridge | Lasso |
| Gaussian | Best Lambda for maximum accuracy: 0.1  Best Validation Accuracy: 100.0 | Best Lambda for maximum accuracy: 0.2  Best Validation Accuracy: 100.0 |
| Random | Best Lambda for maximum accuracy: 0.7  Best Validation Accuracy: 99.27 | Best Lambda for maximum accuracy: 0.7  Best Validation Accuracy: 98.54 |
| Uniform | Best Lambda for maximum accuracy: 0.1  Best Validation Accuracy: 97.81 | Best Lambda for maximum accuracy: 0  Best Validation Accuracy: 97.81 |