**EAS 595 Fundamental of Artificial Intelligence**

**Project#1**

**Dhruv S. Patel(#50321707) Feb-26-2020**

**Task 1**

Results from WEKA

Logistic Regression with ridge parameter of 1.0E-8

**Coefficients**

Variable M

==========================================

radius\_mean -830.3647

texture\_mean 31.7987

perimeter\_mean -27.967

area\_mean 8.7446

smoothness\_mean 8685.1097

compactness\_mean -15608.5869

concavity\_mean 12946.3434

points\_mean 6164.7105

symmetry\_mean -6479.0842

dimension\_mean 18915.9307

radius\_se 2899.827

texture\_se -185.1696

perimeter\_se -363.7952

area\_se 14.9408

smoothness\_se -37182.8004

compactness\_se 21637.8852

concavity\_se -22064.0861

points\_se 97755.9597

symmetry\_se -31736.4434

dimension\_se -176974.0685

radius\_worst 0.5916

texture\_worst 27.9064

perimeter\_worst 26.4228

area\_worst 1.1531

smoothness\_worst 1313.2698

compactness\_worst -1028.4575

concavity\_worst 398.3218

points\_worst 2554.1534

symmetry\_worst 6606.0244

dimension\_worst 15228.5821

Intercept -156.805928

**Summary**

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 110 | **96.4912%** |
| Incorrectly Classified Instance | 4 | 3.5088% |
| Kappa statistic | 0.9288 |  |
| Mean absolute error | 0.0351 |  |
| Root mean squared error | 0.1873 |  |
| Relative absolute error | 7.307% |  |
| Root relative squared error | 37.4551% |  |
| Total Number of Instances | 114 |  |

**Detailed Accuracy by Class**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|  | 0.980 | 0.046 | 0.941 | 0.980 | 0.960 | 0.929 | 0.976 | 0.940 | M |
|  | 0.954 | 0.020 | 0.984 | 0.954 | 0.969 | 0.929 | 0.977 | 0.977 | B |
| Weighted Avg. | 0.965 | 0.031 | 0.966 | 0.965 | 0.965 | 0.929 | 0.961 | 0.961 |  |

**Confusion Matrix**

|  |  |  |
| --- | --- | --- |
|  | TP | TN |
| M | B |
| FP | 48 | 1 |
| FN | 3 | 62 |

* The confusion matrix is very simple. In the first row, for example, it tells you the number of instances classified in your training data as M that you classified as M (that is, 7) and the number that are classified as M that you classified as B. The second row is equivalent for instances classified as B.

**Task 2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-Score | Support |
| B | 0.99 | 0.97 | 0.98 | 69 |
| M | 0.96 | 0.98 | 0.97 | 45 |

* By using SciKit Learn Library, we get accuracy of 97.36%.
* It is 96% correctly classified and it predicts 98% corrected value(True Positive/Negative, False Positive/ Negative).
* There 100 integration(epochs) in LogisticRegression library. So, by using those libraries, we get this confusion matrix and accuracy.

|  |  |  |
| --- | --- | --- |
|  | TP | TN |
| M | B |
| FP | 44 | 1 |
| FN | 2 | 67 |

* Confusion matrix:
* From confusion matrix, 44(TP) targets are Malign and 1(FP) Malign is considered as Benign. And 67(TN) targets are benign and 2(FN) Benign is considered as Malign.

**Task 3**

* Accuracy = 0.9736 = 97.36%
* Precision = 0.9714 = 97.14%
* Recall = 0.9855 = 98.55%
* From above result, the predicted value is **97.36% accurate**. It is **97.14% correctly classified** and it predicts 98.55% corrected value(True Positive/Negative, False Positive/ Negative).
* Confusion Metrix

|  |  |  |
| --- | --- | --- |
|  | TP | TN |
| M | B |
| FP | 43 | 1 |
| FN | 2 | 68 |

* From confusion matrix, 43(TP) targets are Malign and 1(FP) Malign is considered as Benign. And 68(TN) targets are benign and 2(FN) Benign is considered as Malign.
* By changing epoch(integration) and learning rate, accuracy of the model changes. Higher the epochs and lower the learning rate will give more accurate predictions.