Programming Assignment 2 – Applied Machine Learning

# Results

After coding out the algorithms, we first run Logistic Regression on the simplified dataset. We obtained 100% accuracy on the test set. We then tried running the Naïve Bayes code we wrote on the simplified data set, we again obtained 100% accuracy. These results stumped us, as Prof. Natarajan told us that if you get 100% accuracy the result is wrong.

We started scanning the code for bugs, out code seemed alright, so we then we put the datasets in WEKA. Surprisingly, for Naïve Bayes, even WEKA gave 100% accuracy. However, for Logistic Regression, WEKA gave an accuracy of 94.28%.

## Confusion Matrices

Confusion matrix for Naïve Bayes algorithm with Laplace Correction on the simplified dataset using our code.

|  |  |  |
| --- | --- | --- |
|  | 1 | 0 |
| 1 | 14 | 0 |
| 0 | 0 | 21 |

Predicted Result

Expected Result

Confusion matrix for Logistic Regression algorithm on the simplified dataset using our code.

|  |  |  |
| --- | --- | --- |
|  | 1 | 0 |
| 1 | 14 | 0 |
| 0 | 0 | 21 |

Predicted Result

Expected Result

Confusion matrix for Naïve Bayes algorithm with Laplace Correction in WEKA on the simplified dataset.

|  |  |  |
| --- | --- | --- |
|  | 1 | 0 |
| 1 | 14 | 0 |
| 0 | 0 | 21 |

Predicted Result

Expected Result

Confusion matrix for Logistic Regression algorithm in WEKA on the simplified dataset.

|  |  |  |
| --- | --- | --- |
|  | 1 | 0 |
| 1 | 14 | 0 |
| 0 | 2 | 19 |

Predicted Result

Expected Result

## Observations

When running our logistic regression code, we made a peculiar observation. The algorithm converges only when the ƞ value was greater than the threshold by a factor of 1000. No matter the value of ƞ, it has to obey the above condition. For example, when ƞ is 1.0e-8 then threshold should be 1.0e-5 or greater.

In both the algorithms we implemented, we obtained an accuracy of 100%, which was unusual but when we run the algorithms in WEKA on the given training and test dataset we obtained similar results. In WEKA, Naïve Bayes gives an accuracy of 100% and Logistic Regression gives an accuracy of 94.29%. Hence Naïve Bayes works better than Logistic Regression in this case. We do not know why but our speculation is that Naïve Bayes models a probability for the class labels using the feature distribution and Logistic Regression tries to classify the given sample as positive or negative with the probability of 1 or 0. We cannot apply decision theory on Logistic Regression.