# MACHINE LEARNING

MEHAK BERI | Net Id: MXB166430 | HomeWork-2

* **MULTINOMIAL NAÏVE BAYES ALGORITHM**
* I ran the Naïve Bayes Algorithm and have placed the code in different modular functions and written each function’s description along with for easy understanding of code.
* Following are the results obtained by implementing Multinomial naïve bayes algorithm with add-one Laplace smoothing:

|  |  |  |  |
| --- | --- | --- | --- |
|  | DATASET 1 | DATASET 2 | DATASET 3 |
| Accuracy-Ham | 93.10% | 93.81% | 92.76% |
| Accuracy- Spam | 99.23% | 100% | 88.49% |
| Accuracy-Total | 94.76% | 95.83% | 89.68% |

* **MCAP LOGISTIC REGRESSION**
* I ran the MCAP Logistic regression algorithm with L2 regularization. The vocabulary for 70% data consisted of nearly 8000 words and for the complete training data, nearely 10,000 words.
* I created the weight vector for each of those words, that is, I took the presence of absence of a word in a spam/ham document as its feature (unlike Naïve Bayes where I took frequency of the words as features).
* I have placed the code in different modular functions and written each function’s description along with for easy understanding of code.
* I chose the step size= eta as 0.01 as it was the recommended value.
* I ran the program for estimation of weights for 50 iterations because there were 10,000 weights to be estimated, and each weight estimation took nearly 0.5 second. Therefore, it took 10,000/120 = 84 minutes to run one iteration of weight calculation for my computer.
* From my experiments with different values of lambda, using the validation set, I got the following data:
* **PERCEPTRON ALGORITHM**
* I ran the algorithm using number of iterations as a hyperparameter for the validation set. have placed the code in different modular functions and written each function’s description along with for easy understanding of code. I have chosen eta= step size as 0.1 as that was the recommended value in the book.
* I found the following results:
* Testing on validation training set 1 with a chosen value of iterations as 50

Value of iteration: 50

Accuracy on given set is: 93.5251798561151%; Accuracy on ham: 0.9509803921568627 ; Accuracy on spam: 0.8918918918918919

* Testing on validation training set 1 with a chosen value of iterations as 100

Value of iteration: 100

Accuracy on given set is: 93.5251798561151%; Accuracy on ham: 0.9509803921568627 ; Accuracy on spam: 0.8918918918918919

* Testing on validation training set 1 with a chosen value of iterations as 150

Value of iteration: 150

Accuracy on given set is: 93.5251798561151%; Accuracy on ham: 0.9509803921568627 ; Accuracy on spam: 0.8918918918918919

* Testing on validation training set 1 with a chosen value of iterations as 200

Value of iteration: 200

Accuracy on given set is: 93.5251798561151%; Accuracy on ham: 0.9509803921568627 ; Accuracy on spam: 0.8918918918918919

* Testing on validation training set 1 with a chosen value of iterations as 51

Value of iteration: 51

Accuracy on given set is: 94.24460431654677%; Accuracy on ham: 0.9607843137254902 ; Accuracy on spam: 0.8918918918918919

* Testing on validation training set 1 with a chosen value of iterations as 101

Value of iteration: 101

Accuracy on given set is: 94.24460431654677%; Accuracy on ham: 0.9607843137254902 ; Accuracy on spam: 0.8918918918918919

* Testing on validation training set 1 with a chosen value of iterations as 151

Value of iteration: 151

Accuracy on given set is: 94.24460431654677%; Accuracy on ham: 0.9607843137254902 ; Accuracy on spam: 0.8918918918918919

I did similar tests for validation training set 2 and validation training set 3 and decided to set the value of hyperparamter= Number of iterations = 151

The following were my final results:

Testing on full training set 1 with a chosen value of iterations as 151

Value of iteration: 151

Accuracy on given set is: 92.46861924686193%; Accuracy on ham: 0.9080459770114943 ; Accuracy on spam: 0.9692307692307692

Testing on full training set 2 with a chosen value of iterations as 151

Value of iteration: 151

Accuracy on given set is: 75.87719298245614%; Accuracy on ham: 0.6872964169381107 ; Accuracy on spam: 0.9060402684563759

Testing on full training set 3 with a chosen value of iterations as 151

Value of iteration: 151

Accuracy on given set is: 95.39594843462247%; Accuracy on ham: 0.9210526315789473 ; Accuracy on spam: 0.9667519181585678

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_