**Machine Learning**

**Assignment-2 Email Spam Filtering**

Manish Biyani | mxb172930

* Report the accuracy of three data sets available on class webpage.
* Each data set is divided into two sets: the training set and the test set. Data set contain email text format.
* Accuracy is calculated in 3 ways:

1. Naïve Bayes
2. LR
3. Perceptron

* Accuracy measured are as follows:

1. Naïve Bayes

Data-set 1:

1. Ham: 96.74%
2. Spam: 92.61%
3. Total: 94.68%

Data-set 2:

1. Ham: 96.14%
2. Validation: 87.24%
3. Test: 91.83

Data-set 3:

1. Ham: 86.18%
2. Validation: 96.65%
3. Test: 91.42
4. Perceptron

Data-set 1: eta= 0.1

1. On Validation-set

{Iterations - Accuracy } = { 5 - 61.87%,10 – 92.08%, 15 – 96.4% , 20 – 96.40%}

1. On Final test-set

{Iteration- Accuracy} = { 15 – 97.48%}

Data-set 2: eta= 0.1

1. On Validation-set

{Iterations - Accuracy } = { 5 - 57.03%, 10 – 94.81%, 15 – 97.03% , 20 – 97.03%}

1. On Final test-set

{Iteration- Accuracy} = { 15 – 95.83%}

Data-set 3: eta =0.1

1. On Validation-set

{Iterations - Accuracy } = { 5 - 83.22%, 10 – 92.54%, 15 – 93.16% , 20 – 93.16%}

1. On Final test-set

{Iteration- Accuracy} = { 15 – 97.42%}

1. LR

Data-set 1: eta= 0.02

1. On Validation-set

{lambda- Accuracy } = { 0.05 - 86.95%,3 – 85.50%, 7 – 88.40% , 13 – 80.43%}

1. On Final test-set

{Iteration- Accuracy} = { 7 – 92.25%}

Data-set 2: eta= 0.02

1. On Validation-set

{Iterations - Accuracy } = { 0.05 - 94.02%, 3 – 91.79%, 7 – 91.04% , 13 – 70.89%}

1. On Final test-set

{Iteration- Accuracy} = { 0.05 – 93.42%}

Data-set 3: eta=0.02

1. On Validation-set

{Iterations - Accuracy } = { 0.05 - 96.85, 3 – 97.48%, 7 – 97.48% , 13 – 87.42%}

1. On Final test-set

{Iteration- Accuracy} = { 3 – 96.31%}

Q. Can we get some extra time to submit the extra credit question? I’ve 2 exams on 8th as well as all the assignments are due in this week?

References:

* ML by Tom Mitchell
* Various internet sites
* Mitchell's new book chapter (MCAP Logistic Regression algorithm)
* <http://nlp.stanford.edu/IR-book/pdf/13bayes>.