READ ME

The following packages are required for running algoritms.

nltk, sklearn, numpy, pandas, re(regular expression), glob, sys, math, matplotlib, string.

For running code be at the same file address.

For Generating Bag of Word model dataset:

**python BOW.py <D\_value> <e\_value>**

for eg: *python BOW.py Dataset1 enron*

The above command create the Bag of Word model for Dataset 1

So for creating Bag of Word model,

1] For Dataset1 = **D\_value : Dataset1, e\_value: enron**

2] For Dataset2 = **D\_value : Dataset2, e\_value: enron1**

3] For Dataset3 = **D\_value : Dataset3, e\_value: enron4**

*Note:* ***In this assignment I have generated the Bag of Word model for each datasets and stored them in their respective csv file.***

For Generating Bernoulli model dataset:

**python Bernoulli.py <D\_value> <e\_value>**

for eg: *python Bernoulli.py Dataset1 enron*

The above command create the Bernoulli model for Dataset 1

So for creating Bernoulli model,

1] For Dataset1 = **D\_value : Dataset1, e\_value: enron**

2] For Dataset2 = **D\_value : Dataset2, e\_value: enron1**

3] For Dataset3 = **D\_value : Dataset3, e\_value: enron4**

*Note:* ***In this assignment I have generated the Bernoulli model for each datasets and stored them in their respective csv file.***

I] In order to run Multinomial Naive Bayes on the Bag of words model following command is use:

**python Naive.py <D\_value> <e\_value>**

for eg: *python Naive.py Dataset1 enron*

The above command run Multinomial Naive Bayes algorithm on the Bag of words model for Dataset1

So

1] For Dataset1 = **D\_value : Dataset1, e\_value: enron**

2] For Dataset2 = **D\_value : Dataset2, e\_value: enron1**

3] For Dataset3 = **D\_value : Dataset3, e\_value: enron4**

II] In order to run Discrete Naive Bayes on the Bernoulli model following command is use:

**python Discrete.py <D\_value> <e\_value>**

for eg: *python Discrete.py Dataset1 enron*

The above command run Discrete Naive Bayes algorithm on the Bernoulli model for Dataset1

So

1] For Dataset1 = **D\_value : Dataset1, e\_value: enron**

2] For Dataset2 = **D\_value : Dataset2, e\_value: enron1**

3] For Dataset3 = **D\_value : Dataset3, e\_value: enron4**

III] In order to run Logistic Regression on the Bag of words and Bernoulli models following command is use:

**1] For Bag of words model:**

**python Logistic.py <D\_value> <e\_value> <Lambda> <alpha> <Iteration>**

for eg: *python Logistic.py Dataset1 enron 2 0.29 1000*

The above command runs the Logistic Regression algorithm in Dataset1 with lambda, alpha, and iteration values. Also the values specified for lambda, alpha, and iteration in this text is the tuned values. It can changed.

So,

1] For Dataset1 = **D\_value : Dataset1, e\_value: enron, Lambda: 2, alpha: 0.29, Iteration: 1000**

2] For Dataset2 = **D\_value : Dataset2, e\_value: enron1, Lambda: 2, alpha: 0.2, Iteration: 1000**

3] For Dataset3 = **D\_value : Dataset3, e\_value: enron4, Lambda: 3, alpha: 0.1, Iteration: 900**

**2] Bernoulli model:**

**python LogisticB.py <D\_value> <e\_value> <Lambda> <alpha> <Iteration>**

for eg: *python LogisticB.py Dataset1 enron -3 0.1 900*

The above command runs the Logistic Regression algorithm in Dataset1 with lambda, alpha, and iteration values. Also the values specified for lambda, alpha, and iteration in this text is the tuned values. It can changed.

So,

1] For Dataset1 = **D\_value : Dataset1, e\_value: enron, Lambda: -3, alpha: 0.1, Iteration: 900**

2] For Dataset2 = **D\_value : Dataset2, e\_value: enron1, Lambda: 2, alpha: 0.3, Iteration: 1000**

3] For Dataset3 = **D\_value : Dataset3, e\_value: enron4, Lambda: 3, alpha: 0.01, Iteration: 1500**

III] In order to run SGD Classifier:

**1] For Bag of words model:**

**python SGD.py <D\_value> <Iteration>**

for eg: *python SGD.py Dataset1 1000*

The above command runs the Logistic SGD Classifier on Dataset 1 with number of iteration as a parameter. Also the value for number of iteration specified in this text is tuned. It can be change.

So,

1] For Dataset1 = **D\_value : Dataset1, Iteration: 1000**

2] For Dataset2 = **D\_value : Dataset2, Iteration: 1000**

3] For Dataset3 = **D\_value : Dataset3, Iteration: 150**

**2] Bernoulli model:**

**python SGDB.py <D\_value> <Iteration>**

for eg: *python SGDB.py Dataset1 900*

The above command runs the Logistic SGD Classifier on Dataset 1 with number of iteration as a parameter. Also the value for number of iteration specified in this text is tuned. It can be change.

So,

1] For Dataset1 = **D\_value : Dataset1, Iteration: 90**

2] For Dataset2 = **D\_value : Dataset2, Iteration: 150**

3] For Dataset3 = **D\_value : Dataset3, Iteration: 60**