Homework 2
CS 6375.502: Machine Learning
Spring 2018
Naive Bayes, Logistic Regression and Perceptron for Text Classification

Azadeh Samadian

I implement Naive Bayes and Logistic Regression and Perceptron based on the course material with python.
The following tables are the experimental results of these three algorithms.

To extract the vocabulary dictionary, I removed punctuation list: [',', '?', '!', ':', ';', '(', ')', '[', ']', '{', '}'].

For Logistic Regression and Perceptron, I divide training set to 70%(training) and 30%(validation). By using validation set we can find the best lambda and the best iteration and based on that, train the whole data again. I use the model on the test data to find the accuracy.

Naive Bayes:	On dataset2:	Accuracy: 0.949790794979.
Logistic Regression:	On dataSet1: eta= 0.01 Iteration=80 Lambda Range: (0-10)	Best Lambda is: 5. Accuracy: 0.907949790795.
Perceptron:	On dataset3: eta= 0.01 Iteration Range: (0-200)	Best Iteration: 33. Accuracy: 0.922651933702.

ReadMe:

I write this code in python. I used the pycharm IDE in macbook for compiling the code and getting result.
you can copy and paste each of them in the terminal change the file path as an input.

You can use the following sample to run the file.

---

python NaiveBayes.py <training-set> <test-set>

python NaiveBayes.py ./dataSet1/train/ ./dataSet1/test/

---

python LogisticRegression.py <training-set> <test-set>

python LogisticRegression.py /dataSet1/train/ ./dataSet1/test/

---

python perceptron.py <training-set> <test-set>

python perceptron.py ./dataSet1/train/ ./dataSet1/test/
