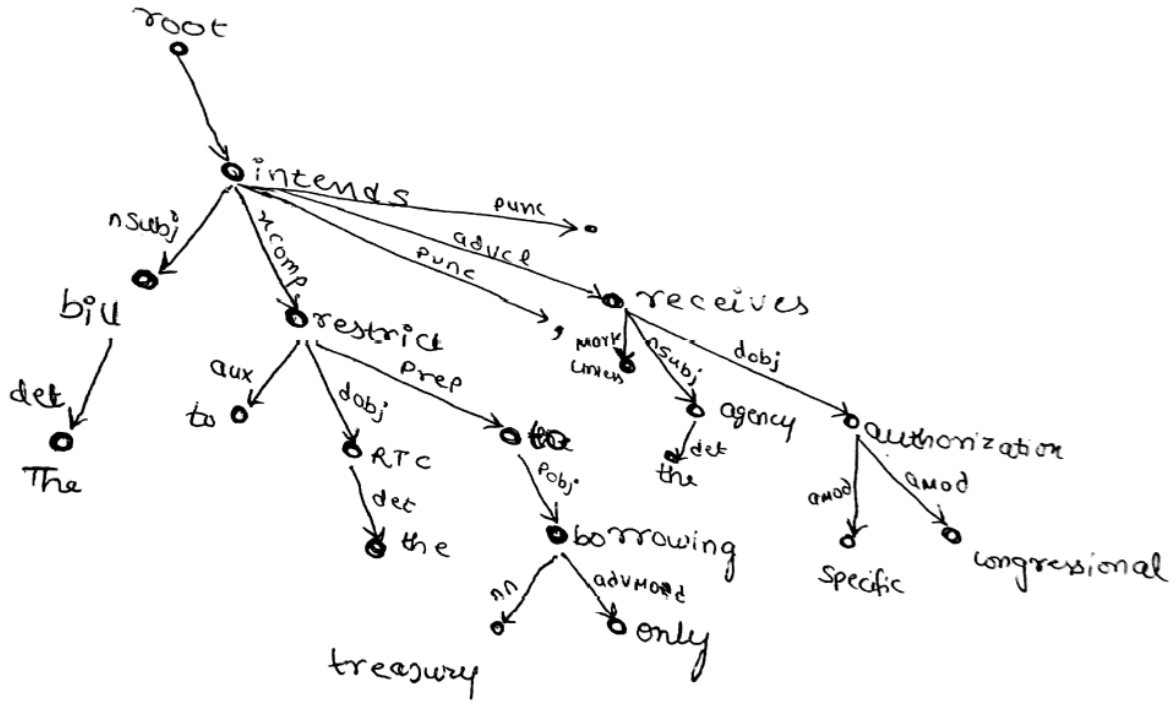
HOMEWORK 4

LING 5801 – Fall 2017

Submitted by
Shivani Sabhlok (500237896)

Question 1 –



Question 2 –

Code submitted on Carmen.

Following is the summary of how the classifier is trained/tested –

- I have reused the ProcessToken method (that removes punctuation marks from the token) from the Zipf's law we implemented previously.
- If the word is a stop word like 'a', 'and', 'the', 'an' ; I have ignored that word because it does not really give any idea about the sentiment.
- After this pre-processing, I have taken three words from the response to train the classifier.
- Similar pre-processing is done of the test data as well.
- To account for the unknown word, I have included smoothing as well while training the classifier.

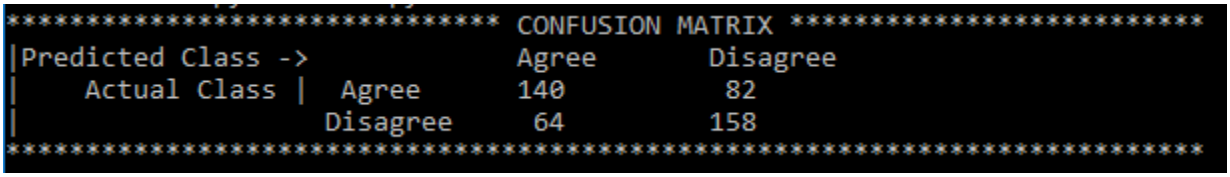
Comments on coding convention mentioned in the code.

Question 3 –

a)

```
Accuracy(percentage) is - 67.117%
Recall(percentage) for Agreement category - 63.063%
Precision(percentage) for Agreement category - 68.627%
Recall(percentage) for Disagreement category - 71.171%
Precision(percentage) for Disagreement category - 65.833%
```

b)



```
********************************** CONFUSION MATRIX ****************************
|Predicted Class ->              Agree       Disagree
|    Actual Class |  Agree       140          82
|                  Disagree      64          158
********************************************************************************
```

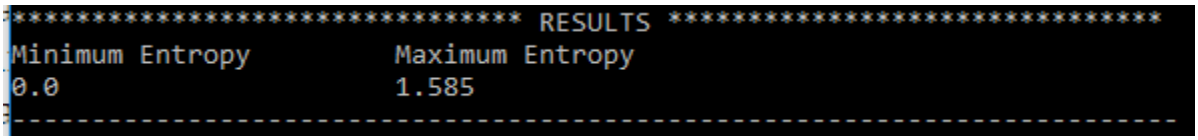
Combined results –

```
C:\Sem1\5801>python HW4.py C:\Sem1\5801\HW4\iac-b-train.csv C:\Sem1\5801\HW4\iac-b-test.csv
****************************** CONFUSION MATRIX ************************
|Predicted Class ->             Agree       Disagree
|    Actual Class |  Agree       140          82
|                  Disagree      64          158
************************************************************************
Accuracy(percentage) is - 67.117%
Recall(percentage) for Agreement category - 63.063%
Precision(percentage) for Agreement category - 68.627%
Recall(percentage) for Disagreement category - 71.171%
Precision(percentage) for Disagreement category - 65.833%
```

Question 4 –

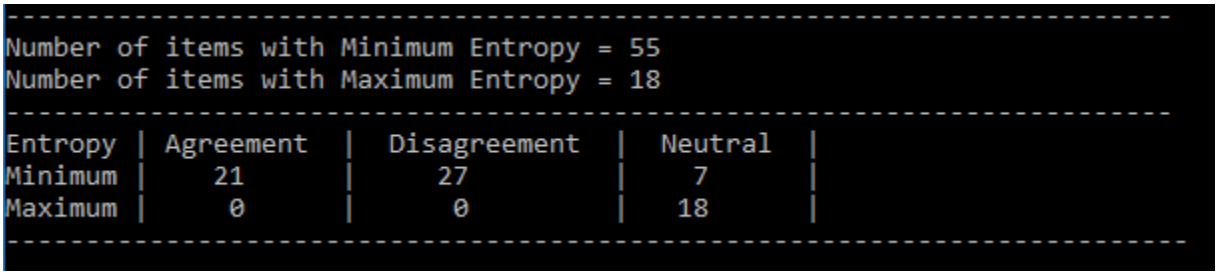
a) Code submitted on carmen.

Results –



```
********************************** RESULTS ***********************************
Minimum Entropy         Maximum Entropy
0.0                     1.585
------------------------------------------------------------------------------
```

b)



```
------------------------------------------------------------------------------
Number of items with Minimum Entropy = 55
Number of items with Maximum Entropy = 18
------------------------------------------------------------------------------
Entropy | Agreement  |  Disagreement  |  Neutral  |
Minimum |    21      |     27         |    7      |
Maximum |     0      |      0         |   18      |
------------------------------------------------------------------------------
```

Combined results –

```
C:\Sem1\5801>python HW4_Entropy.py C:\Sem1\5801\testAnswers.csv
********************************** RESULTS ***********************************
Minimum Entropy         Maximum Entropy
0.0                     1.585
------------------------------------------------------------------------------
Number of items with Minimum Entropy = 55
Number of items with Maximum Entropy = 18
------------------------------------------------------------------------------
Entropy | Agreement  |  Disagreement  |  Neutral  |
Minimum |    21      |     27         |    7      |
Maximum |     0      |      0         |   18      |
------------------------------------------------------------------------------
```