

CS563 – NLP Midsem

Name: P. V. Sriram

Roll No. 1801cs37

Q1)

Run: python Q1.py

With the given train data. I have tested on both test data and P1.txt.
Following is the evaluation metrics in both the cases.

Test Data:

```
1879
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
Accuracy 0.9195359281437125
F score 0.8836508867162285
Precision, Recall, F score (0.8836690792819705, 0.8846892930261586, 0.8836508867162285, None)
```

P1

```
Accuracy 0.9047619047619048
F score 0.9090267928895105
Precision, Recall, F score (0.9171717171717172, 0.9198306113078842, 0.9090267928895105, None)
```

- i) On testing conditions of Para-1 40% of the cases are corrected.
- ii) HMM easily corrects in the case where correct information is there and fails when x is to be detected.

Q2)

Run: python Q2a.py

- a) The following is the tagged version of test input.

```
/* */ That/DT former/JJ Sri/NN Lanka/NNP skipper/NNP and/CC ace/NNP batsman/NNP
Aravinda/NNP De/NNP Silva/NNP is/VBZ a/DT man/NN of/IN few/JJ words/NN was/VBD very/
WRB much/RB evident/VBN on/IN Wednesday/NNP when/WRB the/DT legendary/NNP batsman/
NNP ,/NNP who/WP has/VBZ always/RB let/VBN his/PRP$ bat/NN talk/NN ,/VBD struggled/
CD to/TO answer/CD a/DT barrage/NN of/IN questions/NNS at/IN a/DT function/NNP to_F/
NNPS promote./NNP STOP/WP$
```

Run: python Q2b.py

- b) The tagging will be same as the one in the case above: -

```
/* */ That/DT former/JJ Sri/NN Lanka/NNP skipper/NNP and/CC ace/NNP batsman/NNP
Aravinda/NNP De/NNP Silva/NNP is/VBZ a/DT man/NN of/IN few/JJ words/NN was/VBD very/
WRB much/RB evident/VBN on/IN Wednesday/NNP when/WRB the/DT legendary/NNP batsman/
NNP ,/NNP who/WP has/VBZ always/RB let/VBN his/PRP$ bat/NN talk/NN ,/VBD struggled/
CD to/TO answer/CD a/DT barrage/NN of/IN questions/NNS at/IN a/DT function/NNP to_F/
NNPS promote./NNP STOP/WP$
```

- c) Both the algorithms will result in the same tagging because, even if we consider the top three tags, the tags with lesser probabilities will continue having lesser probabilities in the future steps as well. Since we are considering the total probability to be the product of all previous probabilities, the path with the highest probability will continue till the last. It is important to note that Viterbi already considers all the cases and chooses the most probable path. And output of the next tag depends only on the last two words and their paths. Choosing only the top three paths will not give a better result but the answer will remain the same.