# Home Work 3 - Part 2

Geetanjli Chugh

*University of Florida, United States*

---

---

## 1. Question 1

a)Train a bigram language model on Lincolns speeches (Language model LB)
Approach followed:- Building a Vocabulary :- I am importing input training data depending on different sections i.e various speeches and the building vocabulary using build vocabulary function in nltk.model The first argument to the 'build vocabulary' function is a cutoff value which defines the tokens with frequency counts lesser than mentioned value will be ignored. Here I am using it as 2. In order to count bigrams. I am importing count_ngrams from nltk.model. It will take arguments as count_ngrams i.e highest ngram order to consider, NgramModelVocabulary instance and followed by sequence of sentences respectively. I have used 2 for bigrams, voculabulary bulit in above step and input sentences. While counting ngrams, I am putting the words with very low frequency as "*unknown*. Lapalce model has including smoothing techniques in it. I am using Laplace add-one smoothing which now assigns too much probability to unseen words. I am passing counts to laplace model to calculate perplexity.

I am getting perplexity of 3.6 while testing on lincolns second inaugural speech and using training the bigram model on corpora comprising of Lincolns First Inaugral Address and Gettysburg speech.

Perplexity is 4.85 while testing on nelson speech and using trained bigram model MB test.

Perplexity of LB on LB Train and MB on MB train is 3.54 and 4.72 respectively.

The size of the corpus is very small and hence the size of the vocabulary of both the corpora are very small. There is not much difference in perplexity as we are training model on similar types of text i.e both are speeches. Hence these are not drastically different.

Perplexity of LB on MB Train and MB on MB Train: 3.62, 4.81 respectively. There is slight difference, values have come out to be almost same as part b and d because the speeches are in the same context and the vocabulary of both the speakers are comparable. Hence the perplexity comes out to be approximately same with my approach.

## 2. Question 2

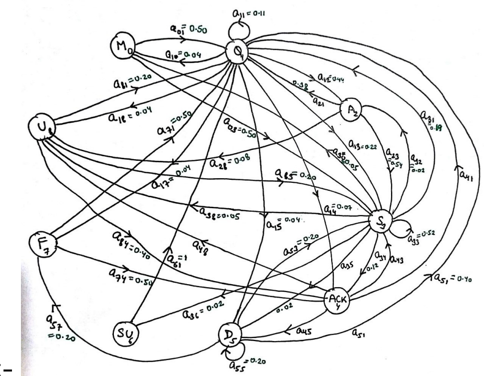
a) I am considering same space as all the tags M,Q,A,S,ACK,D,SU,F,U which are already appearing in the sequence. It can be defined otherwise with all the probable tags with including tags with zero probabilities as well. As we already have the information of tags appearing in the sequence. I am taking tags with non zero probability i.e which are the part of sequence.
S= M,Q,A,S,ACK,D,SU,F,U

Transition Matrix :- A = a01a02 . . .an1 . . .ann a transition probability matrix A, each ai j representing the probability of moving from state I to state j, is as shown below :-

| | M | Q | A | S | ACK | D | SU | F | U |
|---|---|---|---|---|---|---|---|---|---|
| M | 0 | ½ | 0 | ½ | 0 | 0 | 0 | 0 | 0 |
| Q | 1/27 | 3/27 | 12/27 | 6/27 | 2/27 | 1/27 | 0 | 1/27 | 1/27 |
| A | 0 | 5/13 | 0 | 7/13 | 0 | 0 | 0 | 0 | 1/13 |
| S | 2/42 | 8/42 | 1/42 | 22/42 | 5/42 | 1/42 | 1/42 | 0 | 2/42 |
| ACK | 0 | 4/10 | 0 | 3/10 | 0 | 2/10 | 0 | 0 | 1/10 |
| D | 0 | 2/5 | 0 | 1/5 | 0 | 1/5 | 0 | 1/5 | 0 |
| SU | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F | 0 | ½ | 0 | 0 | ½ | 0 | 0 | 0 | 0 |
| U | 0 | 1/5 | 0 | 1/5 | 2/5 | 0 | 0 | 0 | 0 |

Finally Transition matrix after evaluating decimal probabilities is as follows:-

| Column1 | M | Q | A | S | ACK | D | SU | F | U |
|---|---|---|---|---|---|---|---|---|---|
| M | 0.00 | 0.50 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Q | 0.04 | 0.11 | 0.44 | 0.22 | 0.07 | 0.04 | 0.00 | 0.04 | 0.04 |
| A | 0.00 | 0.38 | 0.00 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 |
| S | 0.05 | 0.19 | 0.02 | 0.52 | 0.12 | 0.02 | 0.02 | 0.00 | 0.05 |
| ACK | 0.00 | 0.40 | 0.00 | 0.30 | 0.00 | 0.20 | 0.00 | 0.00 | 0.10 |
| D | 0.00 | 0.40 | 0.00 | 0.20 | 0.00 | 0.20 | 0.00 | 0.20 | 0.00 |
| SU | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| F | 0.00 | 0.50 | 0.00 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 |
| U | 0.00 | 0.20 | 0.00 | 0.20 | 0.40 | 0.00 | 0.00 | 0.00 | 0.00 |

State Diagram for accompanying this model is given below :- 

c) Highest transition probability of all is 1 which occurs when we find probability of Q given SU. P(Q—SU) . Interpretation :- A polite or indirect instruction always leads to question based on given data.

d) In our case minimum achieved probability is 0.02 which is more than cut off value as mentioned in question (0.01). In case we defined a cut off value where we remove lesser transition probabilities. It will remove certain transition links in the state diagram. Removing lesser probabilities will help us in removing the rare cases.