

Report of Assignment 2

Language modelling exercise

Devbrat Anuragi

17078

Q1. Vocabulary:= ['<s>', 'peter', 'piper', 'picked', 'a', 'peck', 'of', 'pickled', 'peppers', '</s>', 'if', 'where', 's', 'the']

Q2 . unigram with highest probability is '<s>' with probability 0.09302325581395349
unigram with second highest Probability 'peter' with probability 0.09302325581395349

Actually <s>, peter,piper, picked,peck,of,picked,peppers,</s> has the same highest probability ,followed by a, if, where, s, the

Q3.

| | <s> | peter | piper | picked | a | peck | of | pickled | peppers | </s> | if | where | s | the |
|---------|-----|-------|-------|--------|------|------|-----|---------|---------|------|------|-------|-----|-----|
| <s> | 0.0 | 0.25 | 0.0 | 0.0 | 0.25 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.25 | 0.25 | 0.0 | 0.0 |
| peter | 0.0 | 0.00 | 1.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| piper | 0.0 | 0.00 | 0.0 | 1.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| picked | 0.0 | 0.00 | 0.0 | 0.0 | 0.50 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.00 | 0.00 | 0.0 | 0.0 |
| a | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| peck | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| of | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| pickled | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| peppers | 0.0 | 0.50 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.5 | 0.00 | 0.00 | 0.0 | 0.0 |
| </s> | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| if | 0.0 | 1.00 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |
| where | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 1.0 | 0.0 |
| s | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 1.0 |
| the | 0.0 | 0.00 | 0.0 | 0.0 | 0.00 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.0 |

Q4.

| | Bigram | Bigram Count | Bigram Prob |
|----|-----------------|--------------|-------------|
| 3 | s the | 1 | 1 |
| 5 | if peter | 1 | 1 |
| 16 | of pickled | 4 | 1 |
| 9 | peter piper | 4 | 1 |
| 1 | peck of | 4 | 1 |
| 12 | pickled peppers | 4 | 1 |
| 13 | the peck | 1 | 1 |
| 14 | where s | 1 | 1 |
| 15 | a peck | 3 | 1 |
| 8 | piper picked | 4 | 1 |

Q5.

[illegible]

Q6. {

```
"</s> </s>": 0.058823529411764705,
"</s> <s>": 0.058823529411764705,
"</s> a": 0.058823529411764705,
"</s> if": 0.058823529411764705,
"</s> of": 0.058823529411764705,
"</s> peck": 0.058823529411764705,
"</s> peppers": 0.058823529411764705,
"</s> peter": 0.058823529411764705,
"</s> picked": 0.058823529411764705,
"</s> pickled": 0.058823529411764705,
"</s> piper": 0.058823529411764705,
"</s> s": 0.058823529411764705,
"</s> the": 0.058823529411764705,
"</s> where": 0.058823529411764705,
"<s> </s>": 0.058823529411764705,
"<s> <s>": 0.058823529411764705,
"<s> a": 0.058823529411764705,
"<s> if": 0.058823529411764705,
"<s> of": 0.058823529411764705,
"<s> peck": 0.058823529411764705,
"<s> peppers": 0.058823529411764705,
"<s> peter": 0.058823529411764705,
"<s> picked": 0.058823529411764705,
"<s> pickled": 0.058823529411764705,
"<s> piper": 0.058823529411764705,
"<s> s": 0.058823529411764705,
"<s> the": 0.058823529411764705,
"<s> where": 0.058823529411764705,
"a </s>": 0.0625,
"a <s>": 0.0625,
"a a": 0.0625,
"a if": 0.0625,
"a of": 0.0625,
"a peppers": 0.0625,
"a peter": 0.0625,
"a picked": 0.0625,
"a pickled": 0.0625,
"a piper": 0.0625,
"a s": 0.0625,
"a the": 0.0625,
"a where": 0.0625,
"if </s>": 0.07142857142857142,
"if <s>": 0.07142857142857142,
"if a": 0.07142857142857142,
"if if": 0.07142857142857142,
"if of": 0.07142857142857142,
"if peck": 0.07142857142857142,
"if peppers": 0.07142857142857142,
"if picked": 0.07142857142857142,
"if pickled": 0.07142857142857142,
"if piper": 0.07142857142857142,
"if s": 0.07142857142857142,
```

"if the": 0.07142857142857142,
"if where": 0.07142857142857142,
"of </s>": 0.058823529411764705,
"of <s>": 0.058823529411764705,
"of a": 0.058823529411764705,
"of if": 0.058823529411764705,
"of of": 0.058823529411764705,
"of peck": 0.058823529411764705,
"of peppers": 0.058823529411764705,
"of peter": 0.058823529411764705,
"of picked": 0.058823529411764705,
"of piper": 0.058823529411764705,
"of s": 0.058823529411764705,
"of the": 0.058823529411764705,
"of where": 0.058823529411764705,
"peck </s>": 0.058823529411764705,
"peck <s>": 0.058823529411764705,
"peck a": 0.058823529411764705,
"peck if": 0.058823529411764705,
"peck peck": 0.058823529411764705,
"peck peppers": 0.058823529411764705,
"peck peter": 0.058823529411764705,
"peck picked": 0.058823529411764705,
"peck pickled": 0.058823529411764705,
"peck piper": 0.058823529411764705,
"peck s": 0.058823529411764705,
"peck the": 0.058823529411764705,
"peck where": 0.058823529411764705,
"peppers </s>": 0.058823529411764705,
"peppers <s>": 0.058823529411764705,
"peppers a": 0.058823529411764705,
"peppers if": 0.058823529411764705,
"peppers of": 0.058823529411764705,
"peppers peck": 0.058823529411764705,
"peppers peppers": 0.058823529411764705,
"peppers picked": 0.058823529411764705,
"peppers pickled": 0.058823529411764705,
"peppers piper": 0.058823529411764705,
"peppers s": 0.058823529411764705,
"peppers the": 0.058823529411764705,
"peppers where": 0.058823529411764705,
"peter </s>": 0.058823529411764705,
"peter <s>": 0.058823529411764705,
"peter a": 0.058823529411764705,
"peter if": 0.058823529411764705,
"peter of": 0.058823529411764705,
"peter peck": 0.058823529411764705,
"peter peppers": 0.058823529411764705,
"peter peter": 0.058823529411764705,
"peter picked": 0.058823529411764705,
"peter pickled": 0.058823529411764705,
"peter s": 0.058823529411764705,
"peter the": 0.058823529411764705,
"peter where": 0.058823529411764705,

"picked </s>": 0.058823529411764705,
"picked <s>": 0.058823529411764705,
"picked if": 0.058823529411764705,
"picked of": 0.058823529411764705,
"picked peck": 0.058823529411764705,
"picked peppers": 0.058823529411764705,
"picked peter": 0.058823529411764705,
"picked picked": 0.058823529411764705,
"picked pickled": 0.058823529411764705,
"picked piper": 0.058823529411764705,
"picked s": 0.058823529411764705,
"picked the": 0.058823529411764705,
"picked where": 0.058823529411764705,
"pickled </s>": 0.058823529411764705,
"pickled <s>": 0.058823529411764705,
"pickled a": 0.058823529411764705,
"pickled if": 0.058823529411764705,
"pickled of": 0.058823529411764705,
"pickled peck": 0.058823529411764705,
"pickled peter": 0.058823529411764705,
"pickled picked": 0.058823529411764705,
"pickled pickled": 0.058823529411764705,
"pickled piper": 0.058823529411764705,
"pickled s": 0.058823529411764705,
"pickled the": 0.058823529411764705,
"pickled where": 0.058823529411764705,
"piper </s>": 0.058823529411764705,
"piper <s>": 0.058823529411764705,
"piper a": 0.058823529411764705,
"piper if": 0.058823529411764705,
"piper of": 0.058823529411764705,
"piper peck": 0.058823529411764705,
"piper peppers": 0.058823529411764705,
"piper peter": 0.058823529411764705,
"piper pickled": 0.058823529411764705,
"piper piper": 0.058823529411764705,
"piper s": 0.058823529411764705,
"piper the": 0.058823529411764705,
"piper where": 0.058823529411764705,
"s </s>": 0.07142857142857142,
"s <s>": 0.07142857142857142,
"s a": 0.07142857142857142,
"s if": 0.07142857142857142,
"s of": 0.07142857142857142,
"s peck": 0.07142857142857142,
"s peppers": 0.07142857142857142,
"s peter": 0.07142857142857142,
"s picked": 0.07142857142857142,
"s pickled": 0.07142857142857142,
"s piper": 0.07142857142857142,
"s s": 0.07142857142857142,
"s where": 0.07142857142857142,
"the </s>": 0.07142857142857142,
"the <s>": 0.07142857142857142,

```

    "the a": 0.07142857142857142,
    "the if": 0.07142857142857142,
    "the of": 0.07142857142857142,
    "the peppers": 0.07142857142857142,
    "the peter": 0.07142857142857142,
    "the picked": 0.07142857142857142,
    "the pickled": 0.07142857142857142,
    "the piper": 0.07142857142857142,
    "the s": 0.07142857142857142,
    "the the": 0.07142857142857142,
    "the where": 0.07142857142857142,
    "where </s>": 0.07142857142857142,
    "where <s>": 0.07142857142857142,
    "where a": 0.07142857142857142,
    "where if": 0.07142857142857142,
    "where of": 0.07142857142857142,
    "where peck": 0.07142857142857142,
    "where peppers": 0.07142857142857142,
    "where peter": 0.07142857142857142,
    "where picked": 0.07142857142857142,
    "where pickled": 0.07142857142857142,
    "where piper": 0.07142857142857142,
    "where the": 0.07142857142857142,
    "where where": 0.07142857142857142
}

```

Q7.

| | c | Nc |
|----------|----------|-----------|
| 0 | 0 | 184 |
| 1 | 1 | 8 |
| 2 | 2 | 4 |
| 3 | 3 | 1 |
| 4 | 4 | 5 |

Q8. We perform smoothing on the count of the counts table in order to account for the unseen words

Q9.

a. MLE Probability = 0.0

b. Laplace Probability = $4.317200740854674e-06$

Bonus: I have tried to do it for the brown corpus but some part is taking very long to complete it took ~ 2 hrs to process 8% of the total bigram. And my system got hanged. Then i tried the same thing on the google colab But there it was giving me the error that your program end up using all the RAM, therefore suspended , So i didn't proceed further.