Assignment 1 Language Model

Chaikesh Chouragade
SR -14358
https://github.com/pkuttam/ngram-assignment

1 Question - Perplexity Measure:

Tr-G = train on gutenberg dataset (80%)
Tr-B = train on brown dataset (80%)
Tr-GB = train on gutenberg and brown dataset (80%)

Ts-G = test on gutenberg dataset (20%)
Ts-B = test on brown dataset (20%)

1.1 Add-K smoothing

parameter k=0.01

—	uni-gram	bi-gram
Tr-B and **Ts-B**	274.9	141.8
Tr-G and **Ts-G**	339	261.8
Tr-GB and **Ts-B**	299	142.6
Tr-GB and **Ts-G**	335	254.7

1.2 kneser-Nay smoothing

d=0.75

—	bi-gram
Tr-B and **Ts-B**	97
Tr-G and **Ts-G**	160
Tr-GB and **Ts-B**	175
Tr-GB and **Ts-G**	110

2 Question - Sentence Generation :

2.1 add-K smoothing

Example- brown dataset i have been a UNK , and the UNK of the UNK , UNK , the UNK . to be a UNK of UNK , a UNK .

Example- Gutenberg dataset

i will not be afraid of the lord , and the lord . make thee a great deal of the house of the children of israel , and he said , ” i am sure i should have been a great many more . give you a great multitude ,

Example- Gutenberg dataset + brown dataset i have not been able to go to the UNK of the lord , and the lord . been a great deal of the UNK , and he said , ” i am sure i should have been the case of the house of the children of israel , and i

[here we can see UNK tag appear due to sparsity in brown dataset]

2.2 kneser-Nay smoothing

Example- brown dataset
i have been , and the UNK of the UNK , and UNK , UNK , the UNK . to be a UNK , a UNK .

Example- Gutenberg dataset

i will not be a great deal of the lord , and the lord . make thee a man of god , and he said , ” i am sure i should be the lord god of israel , and i will give you a great , and to the lord

Example- Gutenberg dataset + brown dataset
i have been a UNK , and the lord , and he said , ” i am sure i should be the lord . not been in the land of egypt , and i will not be a great deal of the lord god of israel , and to the lord

[here we can see UNK tag appear due to sparsity in brown dataset]