

Programming Assignment 1

Machine Learning

-Aaryan Agarwal

Question 1

The first question asked us to evaluate the Maximum Likelihood Estimate, MAP estimate and Predictive Distribution and discuss its effectiveness. This is shown below by calculating the perplexity of each training model and comparing it with test data.

```
The perplexity on train set of MLE model is
size: 640000 Perplexity is 8506.43367662384
size: 160000 Perplexity is 8292.385691215124
size: 40000 Perplexity is 7478.035656314462
size: 10000 Perplexity is 5005.389219343304
size: 5000 Perplexity is 3388.2567752667333
```

```
The perplexity on train set of MAP model is
size: 640000 Perplexity is 8506.96513236839
size: 160000 Perplexity is 8303.124332848962
size: 40000 Perplexity is 7669.433287645091
size: 10000 Perplexity is 6453.994771744834
size: 5000 Perplexity is 5915.104263875246
```

```
The perplexity on train set of Predictive Distribution model is
size: 640000 Perplexity is 8508.427803625034
size: 160000 Perplexity is 8324.246394665119
size: 40000 Perplexity is 7866.496544080013
size: 10000 Perplexity is 7230.294305050776
size: 5000 Perplexity is 7014.415012644821
```

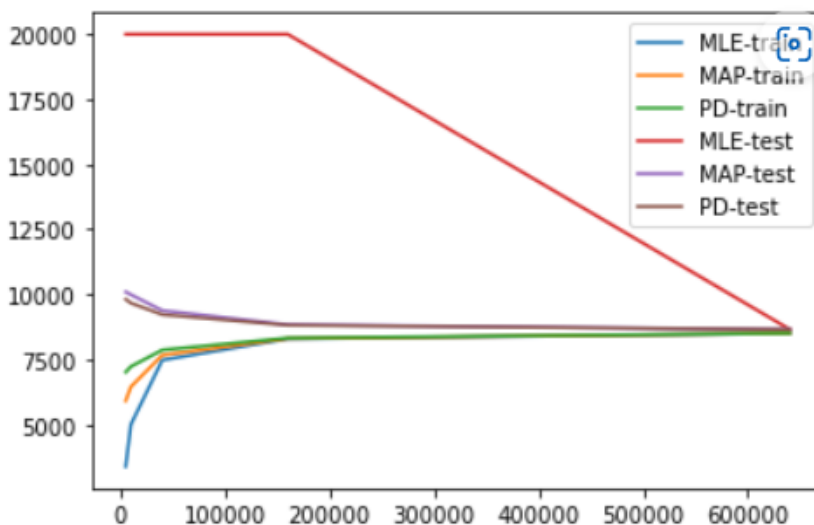
The perplexity on test set of MLE model is
size: 640000 Perplexity is 8657.623041731129

```
C:\Users\Aarya\AppData\Local\Temp\ipykernel_18356\3
p+=np.log(model[i])
```

size: 160000 Perplexity is inf
size: 40000 Perplexity is inf
size: 10000 Perplexity is inf
size: 5000 Perplexity is inf

The perplexity on test set of MAP model is
size: 640000 Perplexity is 8654.590090965366
size: 160000 Perplexity is 8839.546029448937
size: 40000 Perplexity is 9380.752312326787
size: 10000 Perplexity is 9992.362371992125
size: 5000 Perplexity is 10098.36492411617

The perplexity on test set of Predictive Distribution model is
size: 640000 Perplexity is 8652.803792634657
size: 160000 Perplexity is 8817.904839672385
size: 40000 Perplexity is 9224.511912933269
size: 10000 Perplexity is 9668.062580182157
size: 5000 Perplexity is 9814.024919445475



The test set perplexities of the different methods decrease with increase in Test data set size. This occurs because of the increase in prior knowledge and words. The increase in word count results in a

greater data set which helps the model to generate more accurate predictions thus decreasing the perplexity of the model.

The obvious shortcoming of Maximum Likelihood estimation would be that if a word is not present in the training data, then the probability of the word appearing in the test data would be considered as 0. This can result in false results for words which are not present in the training data.

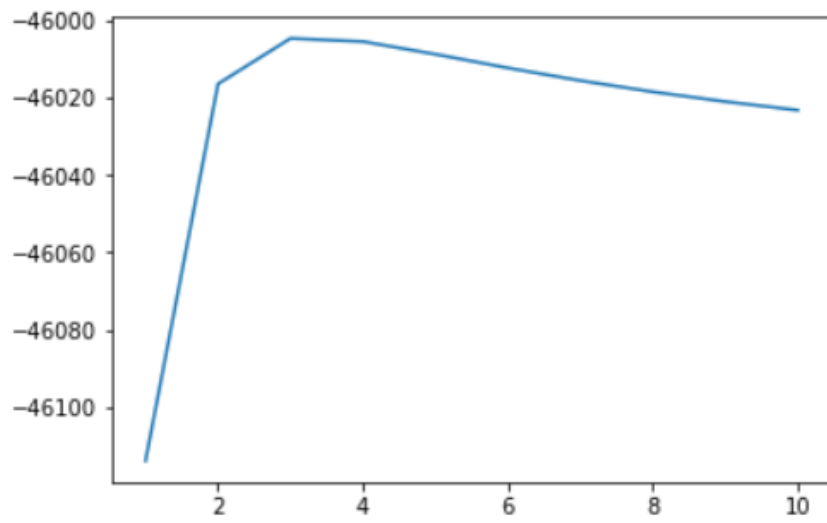
The test set perplexity is not very sensitive to changes in a' because the change in numerator will be negligible with respect to the denominator so the perplexity of the model won't have much affect with the changes.

Question 2

In this question we solve for log evidence and perplexity by using the predictive distribution and the dictionary size $K=10000$ from the Question 1.

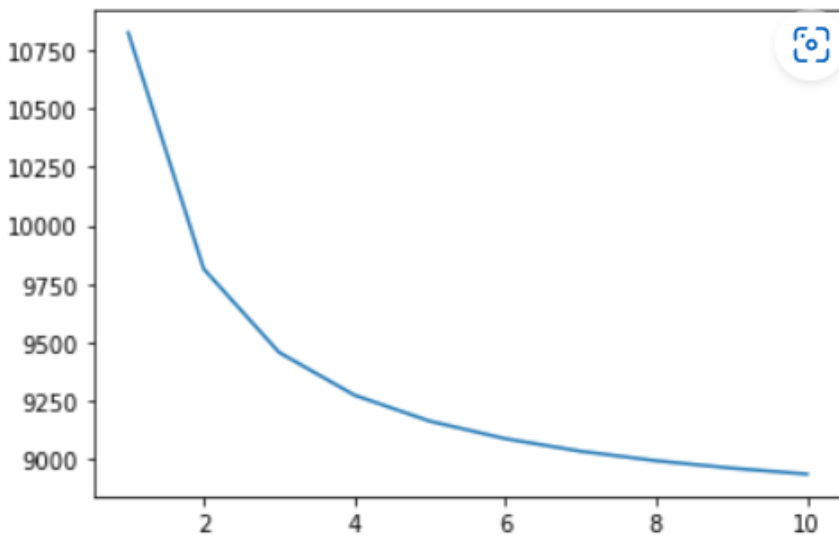
Log evidence:

```
1 -46113.90994393523
2 -46016.4221833731
3 -46004.650537487054
4 -46005.47130750775
5 -46008.75036478405
6 -46012.29823476736
7 -46015.57636918465
8 -46018.47568450381
9 -46021.00839883395
10 -46023.21824466319
```



Perplexity

1 10826.772237368896
2 9814.024919445475
3 9455.9929217656
4 9272.554265719382
5 9160.984335912612
6 9085.95343431994
7 9032.033421256225
8 8991.411806092978
9 8959.708226012419
10 8934.276020578134



Yes, maximizing the evidence function is a good method for model selection as it decreases the perplexity of the model.

Question 3

Perplexity of file 1 3345.6956693728525

Perplexity of file 2 4784.495859529101

Perplexity of file 3 6397.17011622876

File 1 is 121 written by J. Austen

File 2 is 141 written by J. Austen

File 3 is 1400 written by C. Dickens

The perplexity of files 1 and 2 are close to each other so we can cluster them together and conclude that both the files are written by a single author.