Assignment 07, Problem 01

Ankit Agrawal (2581532)

Akshay Joshi (2581346)

Problem 1:

|  |  |  |  |
| --- | --- | --- | --- |
| **epsilon(n)** | **perplexity** | **no. of unigrams** | **no. of bigrams** |
| 3 | 1262.3652 | 119 | 39524 |
| 4 | 545.21457 | 1045 | 50270 |
| 5 | 466.33109 | 11469 | 54359 |
| 6 | 466.33109 | 11469 | 54359 |

As we see, the perplexity decreases with the increase in value of n.

Also, we can see that perplexity is same for epsilon where n=5 and n=6, as nothing gets pruned beyond a certain value of n because no probability is less than that value. So we retain all the ngrams and we don't need to back off to the lower order probability for them.