D3

Alikhon, Jacob, Maksym, Olufela

April 2025

1 What was easy about this assignment?

The assignment consists essentially of the techniques and concepts that were learned during the lectures in the class. Therefore, it was easy to come up with the overall design of the program. The general structure of how the N-gram model should work, the add-1 smoothing function, and the function that computes perplexity was relatively easy to implement due to the previous acknowledgment of the concepts. Once we figured out how to tokenize properly and count unigrams and bigrams, calculating everything else was way easier.

2 What was challenging about this assignment, or parts that you couldn't get working correctly?

Implementing the iterators such as for loops was a little confusing and sometimes challenging. It was hard to come up with the optimal solution and how to efficiently optimize the compilation and execution of the program. For the add-1 smoothing function, for example, our first implementation consisted of the inner for loop, which resulted in a very complicated computation process and eventually the termination of the execution due to unacceptably long computation time. Most of the other challenges and problems were minor and consisted of using the python's dictionary to correctly implement the functions

3 What did you like about this assignment?

We enjoyed seeing how language models work under the hood. It was satisfying to build an NLP system from scratch and see how different aspects impact the predictions and probabilities. The assignment shows that NLP systems are very often implemented differently and there is no one specific implementation or solution to a problem. The assignment also offered a very good chance to connect all concepts learned during lectures to practice.

4 What did you dislike about this assignment?

We mostly liked everything about the assignment. However, sometimes there were moments when debugging and choosing the right way of tokenization were frustrating. It was hard sometimes to understand whether the output was acceptable and within the range of "good" evaluation or not.

5 How did your team function? Include details regarding what each team member contributed, how the team communicated with each other, and how team software development & design was accomplished.

Our team collaborated well and each team member contributed to the final product equally.
Jacob: Handled Unigram and Bigram counting as well as general MLE implementation.
Alikhon: Worked on the function generating sentences and debugging the overall program.
Maksym: Worked on add-1 smoothing and perplexity calculating function.
Olufela: Focused on testing and analysis of different corporas used to test the program.
We communicated primarily over Messages, and used GitHub to share our versions of the code. Everyone messaged regularly and reviewed each other's changes.

6 What did you learn from this assignment?

This assignment helped us to understand how preprocessing affects the overall implementation of the program as well as how it affects the results. Moreover, we learned how smoothing affects the probabilities and how perplexity measures the quality of the model. We also improved our practical knowledge of how to implement NLP models and our skills in working as a team.