EMBEDDED MACHINE LEARNING FOR EDGE COMPUTING

ASSIGNMENT 2 – LET'S BUILD GPT

NADHA IRSHAD SKF2400131

This report compares the performance of different activation functions and tokenization methods in a text generation or character-level prediction model. The focus is on four variations: character-level tokenization with ReLU and GELU activations, and word-level tokenization with ReLU and GELU activations.

Character level tokenization

With Relu : The text generated using ReLU at the character level appears disjointed and contains many nonsensical words and phrases. The structure is fragmented, with frequent abrupt shifts in topic and random insertions of punctuation. Although it occasionally produces recognizable words, the overall coherence is poor, and the text lacks logical flow.

With GELU : The text generated with GeLU at the character level is slightly more coherent than the ReLU output. It still contains many nonsensical phrases and incomplete thoughts, but the transitions between words and phrases are somewhat smoother. The text exhibits better structure, although it still lacks logical sense.

Word level tokenization

With Relu : The word-level text generated using ReLU is more structured and coherent compared to the character-level output. It follows a more logical sequence of words, with phrases that resemble actual sentences and even some semblance of dialogue. However, the content still lacks overall meaning, with many sentences being incomplete or out of context.

With GELU : Provided smoother and more contextually relevant sequences with less repetition and a more natural language flow. This is especially noticeable in how GELU handled transitions between dialogue lines, resulting in text that feels more naturally structured and contextually consistent. It significantly improved the coherence and fluency of word-level predictions. The words generated under GELU activation tend to form more contextually appropriate phrases, maintaining a better flow and reducing the likelihood of producing random or unrelated words.

Conclusion

The text at the character level is more challenging for the model to generate meaningful content. Both ReLU and GeLU produced text with fragmented and nonsensical content, though GeLU was slightly better in terms of flow. The word-level generation, resulted in more structured and coherent output. The text resembles real sentences and dialogue. Overall, word-level text generation yields more coherent and readable text, whereas character-level generation, while producing interesting patterns, often results in gibberish or highly fragmented content.

The GELU based outputs are more contextually appropriate and fluent, offering a more natural dialogue structure and making it a preferred choice for generating text with a higher level of coherence and readability. This improvement is consistent across both **character level** and **word level** outputs, demonstrating the benefits of using **GELU** over **ReLU** in the model architecture.

References: https://www.baeldung.com/cs/gelu-activation-function