November 2025

Assignment III - Exploring Transformer models

Introduction

In this exercise, you will explore the world of Transformers and gain hands-on experience with popular models, including GPT-2 from Hugging Face. You will explore the model's structure, attention mechanism, and gain insights into attention scores. Additionally, you will generate data using diverse strategies with GPT-2

1. Setting up the Environment:

- Install the required Python libraries, including *transformers* from Hugging Face.
- Download a pre-trained GPT-2 model from Hugging Face's model hub.
- Initialize one of the model versions and tokenizer for further exploration.

2. Understanding the GPT-2 Model:

- Investigate the architecture of the GPT-2 model. What are the key components, layers, and parameters involved?
- Apart from the GPT model, the Hugging Face Transformers library has other model architectures such as BERT, BART or T5, how do the model architectures differ? How are the attention mechanisms implemented in these cases? What are the advantages and disadvantages?

3. Generating Data through Diverse Strategies:

- Generate text using the GPT-2 model with different strategies, such as:
- Prompts: Provide a prompt and observe the model's generation based on it.
- Temperature control: Experiment with different temperature values to control the randomness of generated text.
- Top-k sampling: Generate text by selecting from the top-k most likely tokens.
- Top-p sampling (nucleus sampling): Generate text by selecting from the top-p (nucleus) probability distribution.

4. Analysis and evaluation

- Analyze the generated text from different strategies. What are the strengths and weaknesses of each approach?
- Compare and contrast the output generated by GPT-2 under various conditions

5. Deep-dive into attention scores:

- Analyze attention scores produced by the GPT-2 model. What do these scores represent, and how can they be interpreted?
- Visualize attention matrices for specific input sequences. Observe how attention is distributed across tokens

6. [Extra] Fine-tune GPT2 model

Using the movie scripts dataset available in Kaggle (or any other one dataset of your choice). Fine-tune the GPT model for a couple of epochs. Include a prompt and evaluate how the text generated from a fine-tuned model differs from the previous text generation.

Bibliography:

Hugging Face Transformers Library: https://huggingface.co/transformers/

GPT-2 Model from Hugging Face Model Hub: https://huggingface.co/models