ASSIGNMENT - 7

1. Explain the architecture of BERT

Ans: BERT (Bidirectional Encoder Representations from Transformers) leverages a multilayer bidirectional transformer encoder. This means it builds upon the Transformer architecture, focusing solely on the encoder portion (as it's not for generation). Here's a breakdown:

* Transformer Encoder: This is the core building block, using a self-attention mechanism to analyze the relationships between words in a sentence, considering both preceding and following context. This allows BERT to capture rich contextual representations of words.
* Multilayer: BERT stacks multiple encoder layers on top of each other. Each layer refines the understanding of the sentence based on the previous layer's output.
* Bidirectional: Unlike some language models that only process text left-to-right, BERT considers both directions, enabling a more comprehensive grasp of meaning.

2. Explain Masked Language Modeling (MLM)

Ans: MLM is a pre-training technique used in BERT. It involves:

* Randomly masking out some words (replacing them with a special mask token) in the input text.
* Training the model to predict the masked words based on the surrounding context. This forces BERT to learn deep contextual representations of words, as it needs to understand the surrounding words to predict the masked ones accurately.

3. Explain Next Sentence Prediction (NSP)

Ans: NSP works as follows:

The model is presented with two sentences (A and B).

It needs to predict whether the second sentence (B) logically follows the first sentence (A). This trains BERT to understand relationships between sentences and how they flow together coherently.

4. What is Matthews evaluation?

Ans: Matthews evaluation isn't a specific technique used in BERT, but rather a general evaluation metric for classification tasks. It considers true positives, true negatives, false positives, and false negatives to provide a more balanced assessment of model performance compared to simpler metrics like accuracy.

5. What is Matthews Correlation Coefficient (MCC)?

Ans: This is the numerical measure derived from the Matthews evaluation. MCC ranges from -1 (worst) to +1 (best), indicating the quality of binary (two-class) classification. It takes into account true and false positives/negatives, offering a robust measure that considers both positive and negative class predictions.

6. Explain Semantic Role Labeling

Ans: This is a natural language processing (NLP) task where the model identifies the semantic roles of words within a sentence. For example, in "The dog chased the cat," the dog is the subject (agent) and the cat is the object (patient). BERT can be fine-tuned for semantic role labeling tasks by adapting its output layer to predict these roles for each word in a sentence.

7. Why Fine-tuning a BERT model takes less time than pretraining

Ans: Pre-training BERT is computationally expensive, as it involves training on a massive dataset for a long time. This pre-training phase equips BERT with a strong understanding of general language patterns. Fine-tuning, on the other hand, takes a pre-trained BERT model and adapts it to a specific NLP task by adjusting the final layers of the network. This fine-tuning process is significantly faster because the core language understanding is already captured in the pre-trained model.

8. Recognizing Textual Entailment (RTE)

Ans: RTE is an NLP task where the model determines whether the meaning of one sentence (hypothesis) is entailed by the meaning of another sentence (text). For instance, if the text is "The sun is shining," and the hypothesis is "It is daytime," then RTE would classify this as entailment because the first sentence implies the second. BERT can be fine-tuned for RTE tasks by training it on labeled datasets of entailment relationships.

9. Explain the decoder stack of GPT models.

Ans: Generative Pre-trained Transformer (GPT) models like GPT-3 utilize a decoder stack. The decoder builds upon the encoded representation (often generated by a separate encoder or obtained from another model) to sequentially predict the next word in a sequence. This allows GPT models to generate human-quality text, translate languages, write different kinds of creative content, and more.