**ADITYA AMIN**

**ASSIGN: NLP-07**

1. Explain the architecture of BERT.

Input Representation: BERT takes variable-length sequences of tokens as input. Before feeding the tokens into the model, they are first tokenized into subword units using WordPiece tokenization

Transformer Encoder Blocks: BERT is composed of a stack of identical Transformer encoder blocks. Each block consists of multiple self-attention heads and position-wise feed-forward layers.

Self-Attention: Self-attention is a key component of the Transformer model. In BERT, self-attention allows each token to attend to all other tokens in the input sequence, capturing dependencies and relationships..

Pre-training Objectives: BERT is pre-trained using two unsupervised tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP

Fine-tuning for Downstream Tasks: After pre-training, BERT can be fine-tuned for specific downstream tasks such as text classification, named entity recognition, question answering, and more. Explain Masked Language Modeling (MLM).

1. Explain Next Sentence Prediction (NSP).

Next Sentence Prediction (NSP) is one of the pre-training objectives used in the BERT (Bidirectional Encoder Representations from Transformers) model. It is designed to capture the understanding of the relationship between two consecutive sentences in a given text corpus.

The goal of NSP is to train BERT to predict whether a pair of sentences appears consecutively in the original text or not. It helps BERT learn the contextual relationship between sentences and enables the model to understand the coherence and flow of language.

1. What is Matthews evaluation?

Matthews evaluation refers to the Matthews Correlation Coefficient (MCC), which is a measure used to assess the quality of binary (two-class) classification models. It takes into account true positives, true negatives, false positives, and false negatives to provide a balanced evaluation of a classifier's performance, especially when dealing with imbalanced datasets.

The formula for calculating the Matthews Correlation Coefficient is as follows:

MCC = (TP × TN - FP × FN) / sqrt((TP + FP) × (TP + FN) × (TN + FP) × (TN + FN))

1. What is Matthews Correlation Coefficient (MCC)?

The Matthews Correlation Coefficient (MCC) is a measure used to evaluate the performance of binary (two-class) classification models. It takes into account true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) to provide a balanced assessment of the classifier's performance, especially in cases of imbalanced datasets.

1. Explain Semantic Role Labeling

Semantic Role Labeling (SRL) is a natural language processing (NLP) task that aims to identify and classify the semantic roles played by different constituents (words or phrases) within a sentence. It involves understanding the relationships between different words and their roles in conveying the meaning of a sentence.

The goal of SRL is to assign specific labels or roles to words that indicate their semantic function in a sentence. These roles typically capture the predicate-argument structure of the sentence, where the predicate is the main verb or action word, and the arguments are the words or phrases that participate in the action.

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

Fine-tuning a BERT model typically takes less time than pretraining because pretraining involves a significantly larger amount of data and computational resources.

During pretraining, a BERT model is trained on a large corpus of unlabeled text data, such as Wikipedia or web documents. This process involves predicting masked words within the text and determining the relationship between pairs of sentences. Pretraining requires a massive amount of compute power, specialized hardware (e.g., GPUs or TPUs), and substantial time to process the vast amount of data.

1. Recognizing Textual Entailment (RTE)

Recognizing Textual Entailment (RTE) is a natural language processing (NLP) task that involves determining the logical relationship between two pieces of text: a "text" sentence and a "hypothesis" sentence. The goal is to determine whether the meaning of the hypothesis can be inferred or logically derived from the meaning of the text.

1. Explain the decoder stack of GPT models.

The decoder stack in GPT (Generative Pre-trained Transformer) models refers to the series of transformer decoder layers used in the model architecture. GPT models are based on the Transformer architecture and are primarily designed for generating text and language understanding tasks.