**Assignment - 7**

1. Explain the architecture of BERT

Ans: BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based neural network architecture designed for natural language understanding tasks. It consists of multiple layers of self-attention mechanisms and feedforward neural networks. BERT employs a bidirectional approach, allowing it to capture context from both left and right directions in a sequence simultaneously. It consists of an encoder architecture composed of multiple transformer blocks, with each block containing self-attention layers and position-wise feedforward layers. BERT uses a pretraining phase followed by fine-tuning for specific downstream tasks, achieving state-of-the-art performance on various natural language processing tasks.

1. Explain Masked Language Modeling (MLM)

Ans: Masked Language Modeling (MLM) is a pretraining objective used in BERT where a certain percentage of input tokens in a sentence are randomly masked, and the model is trained to predict the masked tokens based on the surrounding context. By masking tokens and predicting them, BERT learns contextualized representations of words that capture bidirectional dependencies in the input text.

1. Explain Next Sentence Prediction (NSP)

Ans: Next Sentence Prediction (NSP) is another pretraining objective used in BERT, where the model is trained to predict whether a pair of sentences in a document appear consecutively or not. NSP helps BERT learn relationships between sentences and improve its understanding of document-level semantics.

1. What is Matthews evaluation?

Ans: Matthews evaluation is a metric used to evaluate the performance of binary classification models, particularly when dealing with imbalanced datasets. It takes into account true positives, true negatives, false positives, and false negatives to compute a correlation coefficient that ranges from -1 to +1, where +1 indicates perfect classification, 0 indicates random guessing, and -1 indicates total disagreement between predictions and ground truth.

1. What is Matthews Correlation Coefficient (MCC)?

Ans: Matthews Correlation Coefficient (MCC) is a measure of the quality of binary (two-class) classifications. It takes into account true positives, true negatives, false positives, and false negatives and is calculated using the formula:

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

MCC ranges from -1 to +1, where +1 indicates perfect prediction, 0 indicates random prediction, and -1 indicates total disagreement between predictions and ground truth.

1. Explain Semantic Role Labeling

Ans: Semantic Role Labeling (SRL) is a natural language processing task where the goal is to identify the semantic roles of words or phrases in a sentence and assign them labels such as "agent," "patient," "instrument," etc. SRL helps in understanding the relationships between different entities and actions in a sentence.

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

Ans: Fine-tuning a BERT model takes less time than pretraining because the pretraining phase involves training the model on a large corpus of text data to learn general language representations, while fine-tuning involves adapting the pretrained model to specific downstream tasks by training it on task-specific datasets with fewer examples. Since fine-tuning requires fewer training steps and less data, it typically takes less time compared to the pretraining phase.

1. Recognizing Textual Entailment (RTE)

Ans: Recognizing Textual Entailment (RTE) is a natural language understanding task where the goal is to determine whether a given piece of text (the hypothesis) logically follows from another piece of text (the premise). RTE is often formulated as a binary classification task, where the model predicts whether the relationship between the premise and hypothesis is entailment, contradiction, or neutral.

1. Explain the decoder stack of GPT models.

Ans: The decoder stack of GPT (Generative Pre-trained Transformer) models consists of multiple layers of decoder blocks. Each decoder block contains self-attention layers and feedforward neural networks similar to those in the encoder stack. During generation, the decoder stack generates output tokens autoregressively, conditioning on the context tokens generated so far. GPT models are unidirectional language models that generate text in a left-to-right manner, using the transformer decoder architecture.