Bert Analysis

Bert stands for Bidirectional Encoder Representations from Transformers which in simpler terms is a model that has multiple transformer encoding layers that are stacked upon each other. Bert is also bidirectional normally this would mean it would have a forward and a backward pass but in the context of nlp it means Bert considers both the left and right contexts of the word while embedding it allowing for better understanding.

Bert can also be used for transfer learning as it is pretrained on large sizes of text data allowing it to be later fine-tuned for job specific downstream tasks like sentiment analysis, text generation etc. The bidirectionality of Bert allowed it to understand the contextuality of words much better making it less prone to errors in languages

Model Architecture:

- Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). MLM randomly masks words in the input text, and the model predicts the masked words based on the surrounding context. NSP predicts whether a pair of sentences are consecutive in a document to capture both the meaning of individual words and their relationships within a sentence.
- As mentioned before Bert has an multi layered transformer encoder which uses and self-attention mechanism allowing Bert to capture long range dependencies
- Finally, and output layer is to be added for fine tuning for downstream tasks

Finally, Bert is tested and achieves good results on multiple nlp tasks thus improving benchmarks on GLUE,MultiNLi etc.