Introduction to pre trained models: BERT and GPT

27 October 2023 11:27 AM

Bi-directional Encoder Representation of Transformers (BERT)

1.) Pre-train BERT to understand language

2.) Fine tune BERT to learn specific tasks.

Pre-training

Here BERT trains on two supervised tasks, to learn about the language that it will be working with:

1.) Masked language Modelling

→ We put a mask on few words in the sentences and then BERT tries to predict those masked words.

Eg The _____ brown fox _____ jumps over the lazy dog.
mask-1 mask-2

↓ quick ↓ Jumps

BERT will try to produce these words in output.

2.) Next Sentence Predictions

↳ The model takes in two sentences as input, and now it will try to identify whether the second sentence logically follows the first

Fine tuning

Here the BERT is trained further on specific NLP tasks that we want to implement.

Generative Pre-trained Transformers (GPT)

Quite similar to BERT, but the difference is that GPT uses the Decoder of the Transformer Architecture