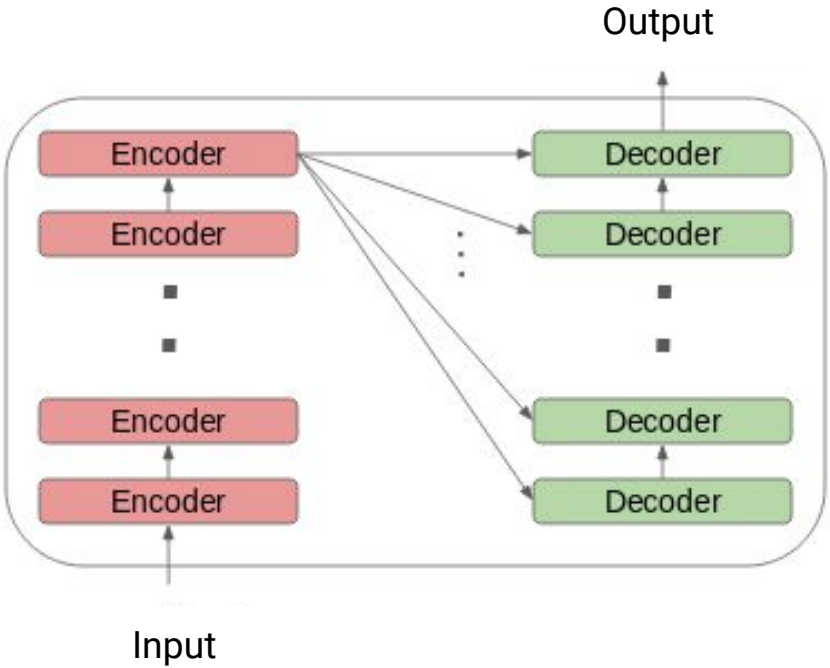
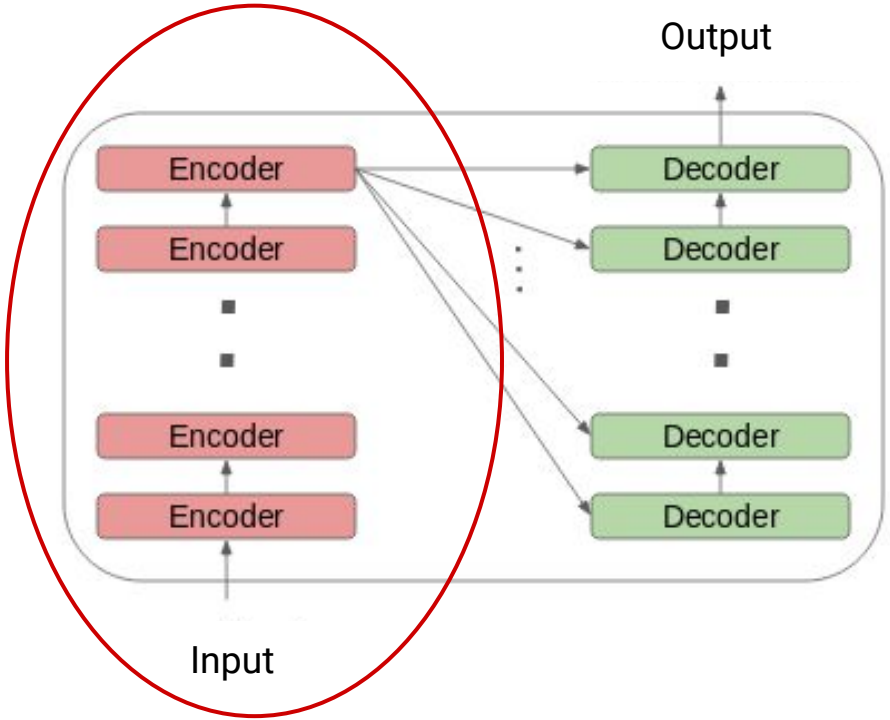
Architecture of BERT

Transformer



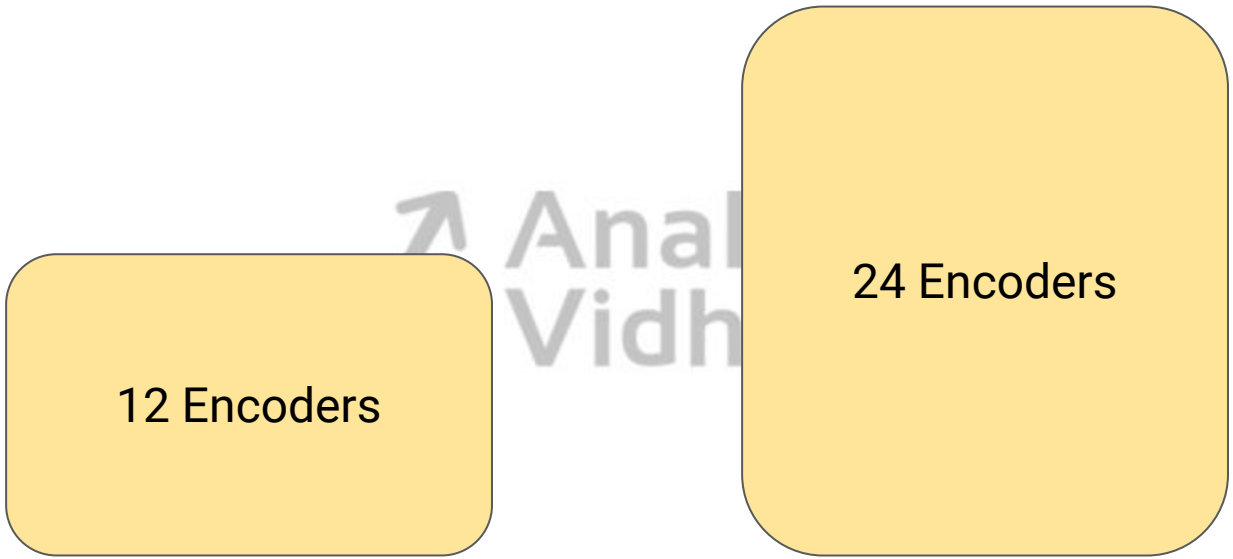
Analytics Vidhya

Transformer



Analytics Vidhya

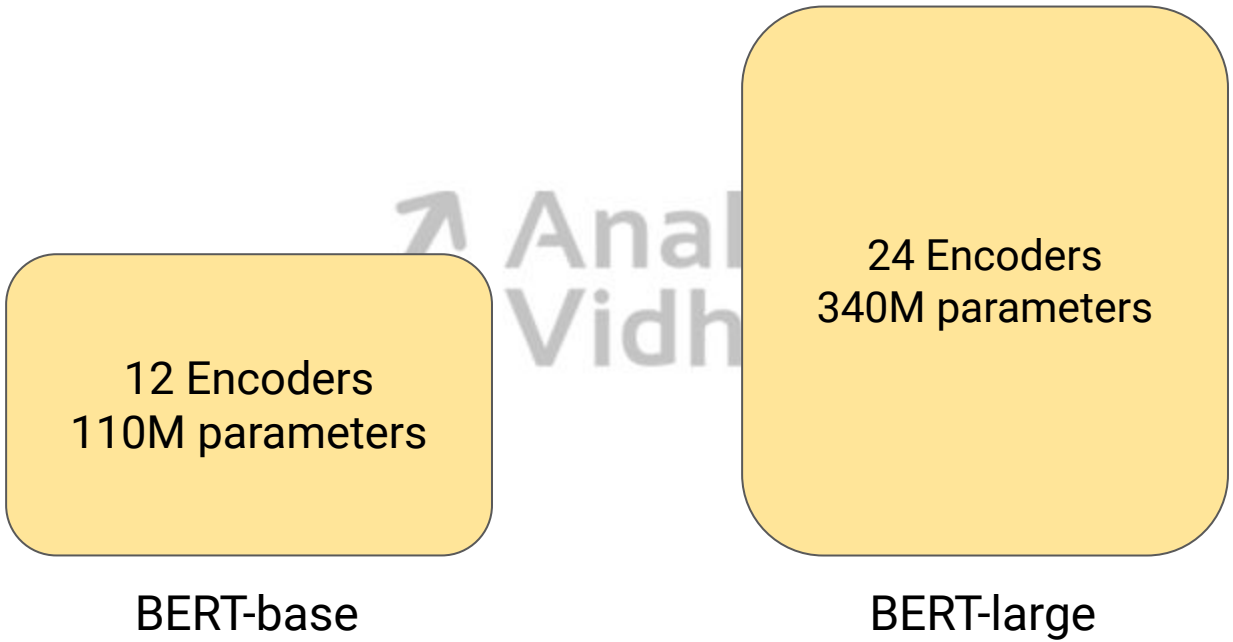
Variants of BERT



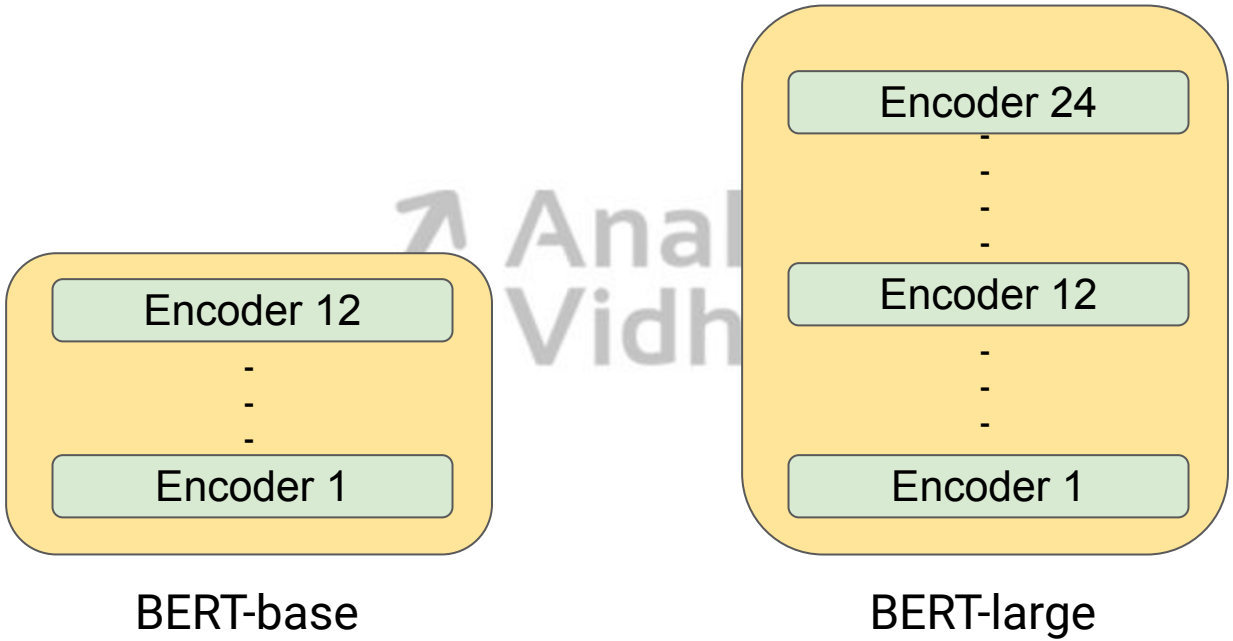
BERT-base

BERT-large

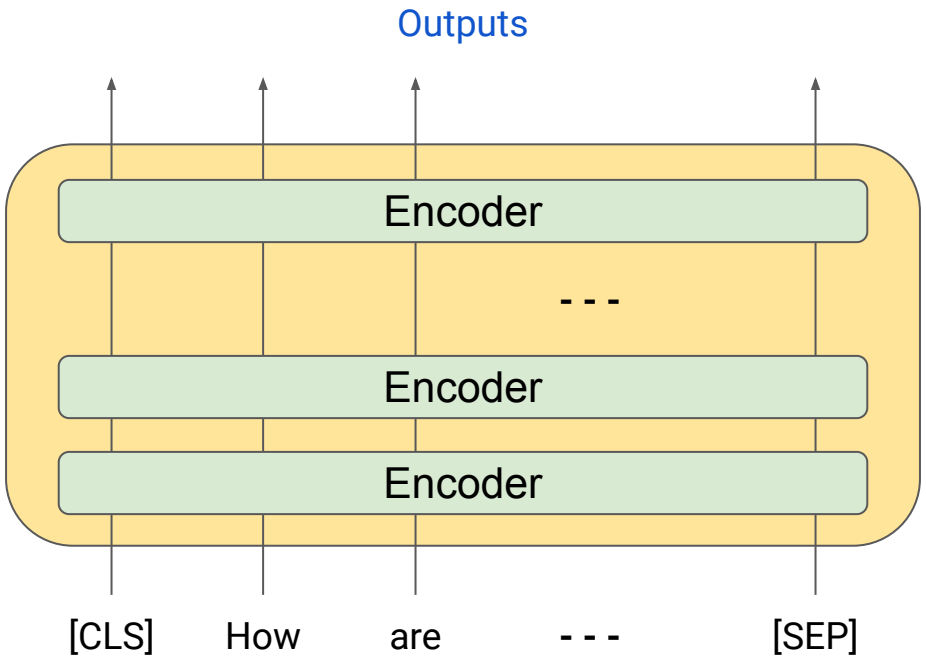
Variants of BERT



Variants of BERT

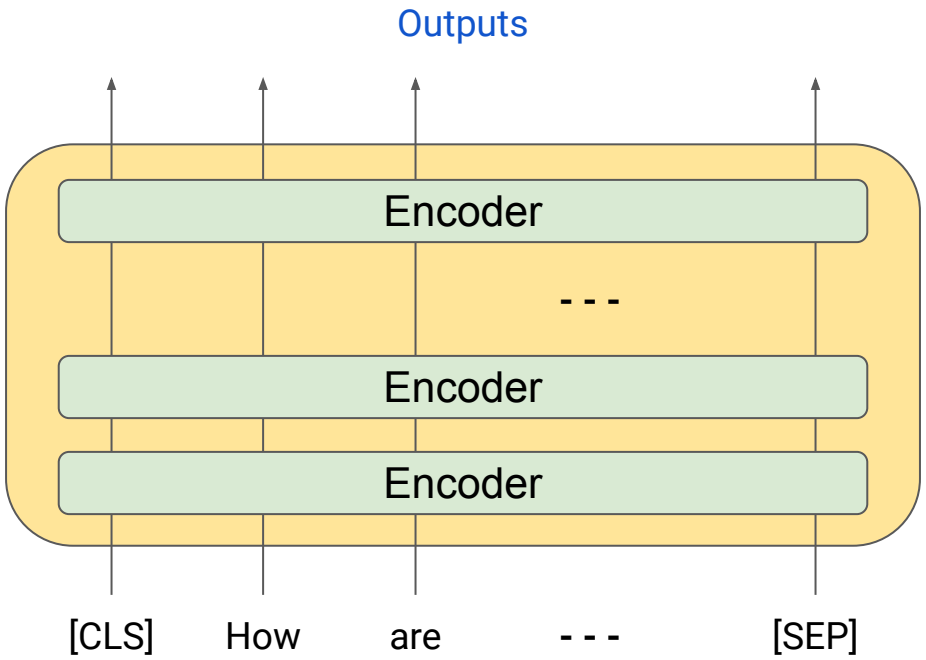


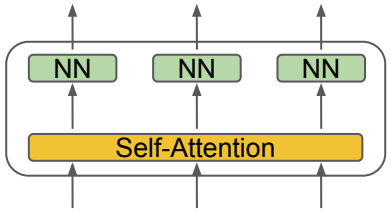
Architecture of BERT



Analytics Vidhya

Architecture of BERT



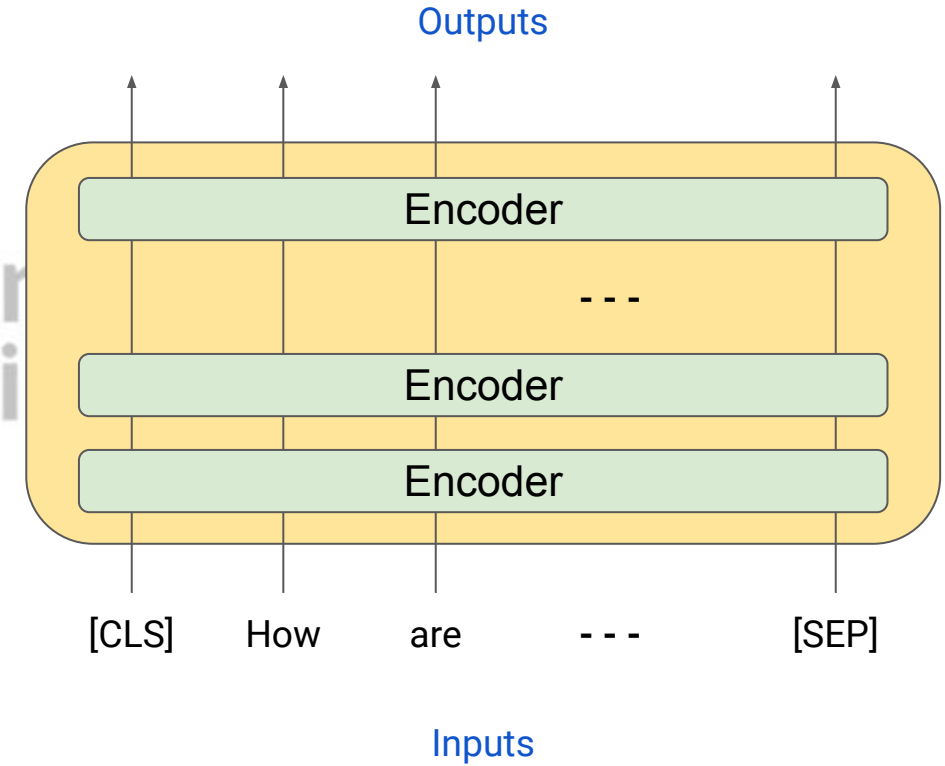


Encoder

Analytics Vidhya

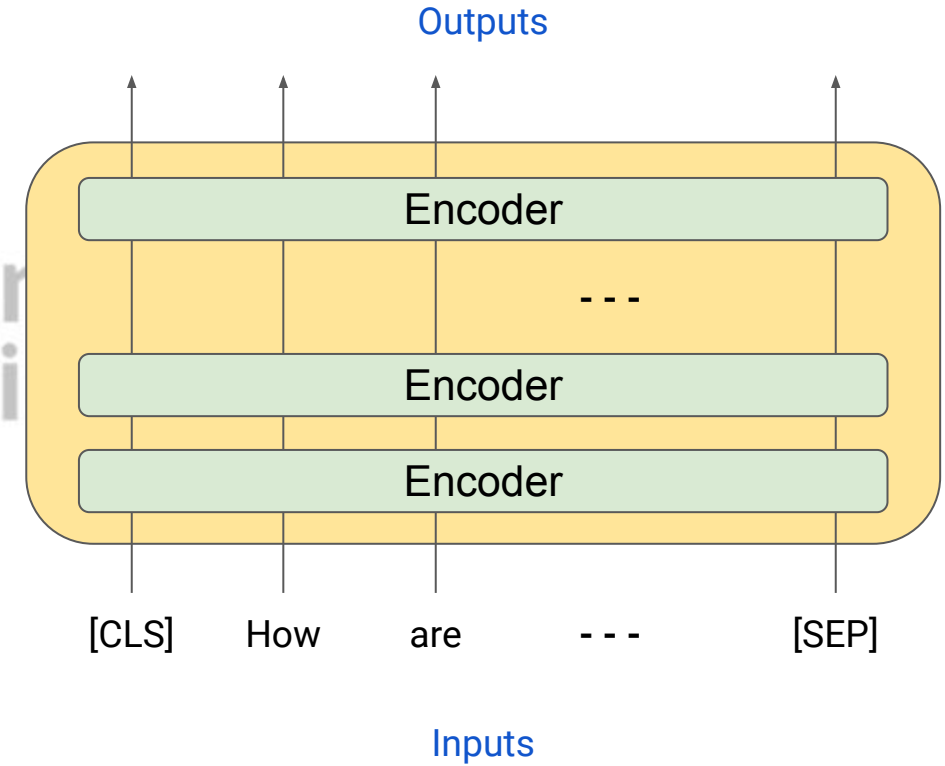
Architecture of BERT

- [CLS] and [SEP] tokens are special tokens.
- [CLS] token is prepended at start and [SEP] token is appended at the end of the sequence



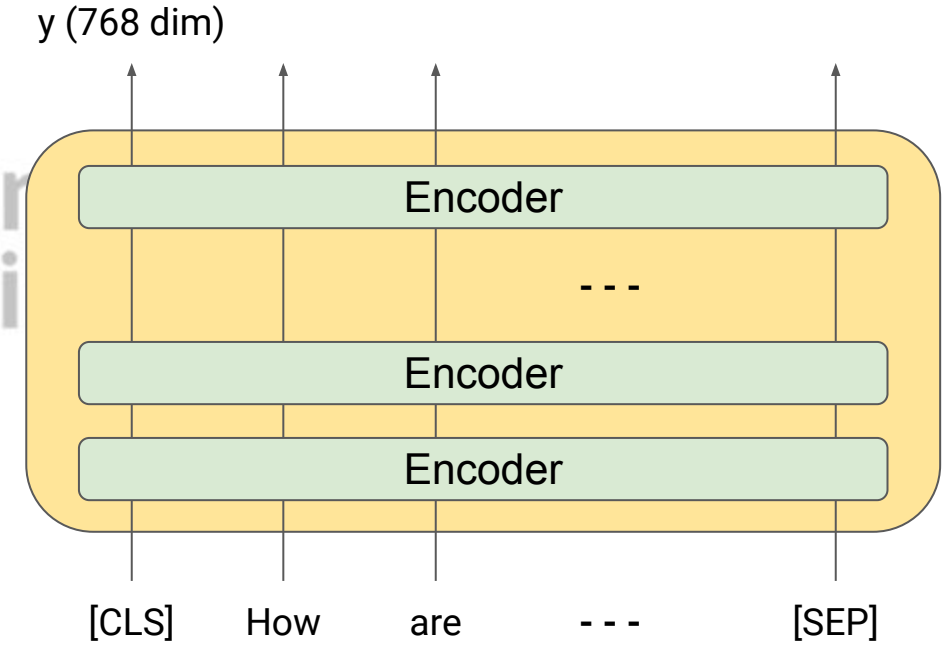
Architecture of BERT

- [CLS] and [SEP] tokens are special tokens.
- [CLS] token is prepended at start and [SEP] token is appended at the end of the sequence
- [CLS] represents the entire sequence

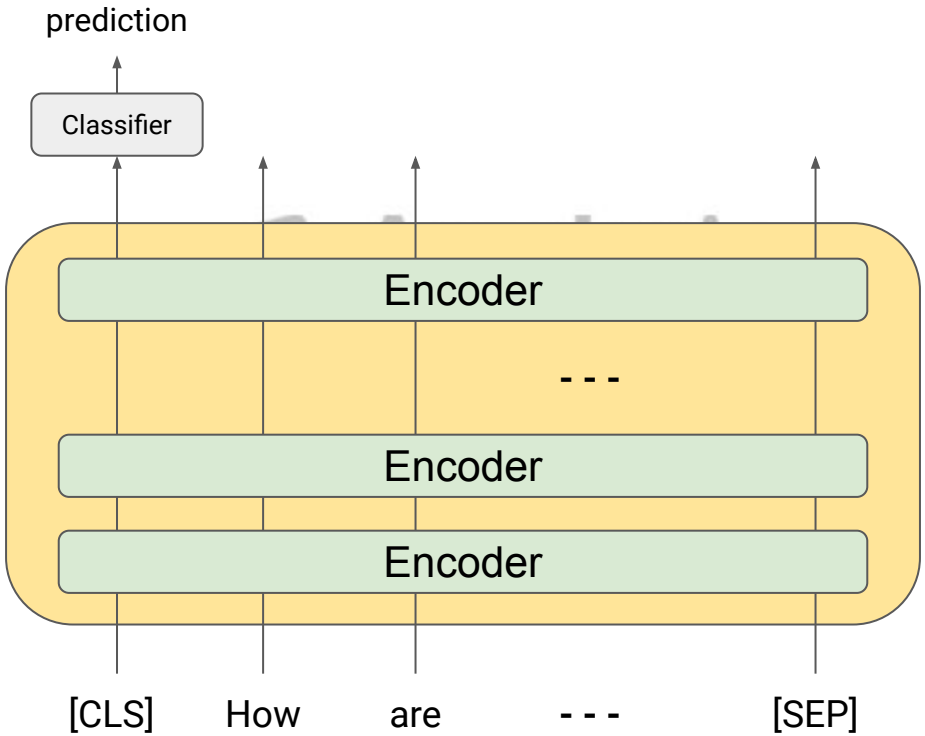


How to Use BERT?

- Feature Extraction
  - Convert input text to vector representation
- Pass the output to an extra linear layer and apply softmax

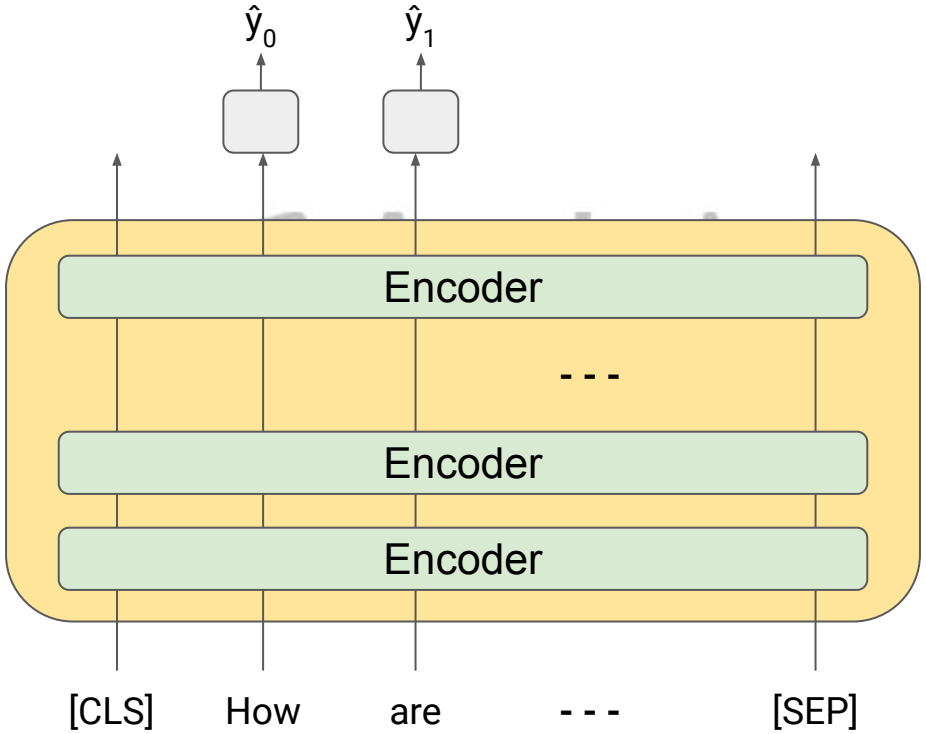


BERT for Text Classification

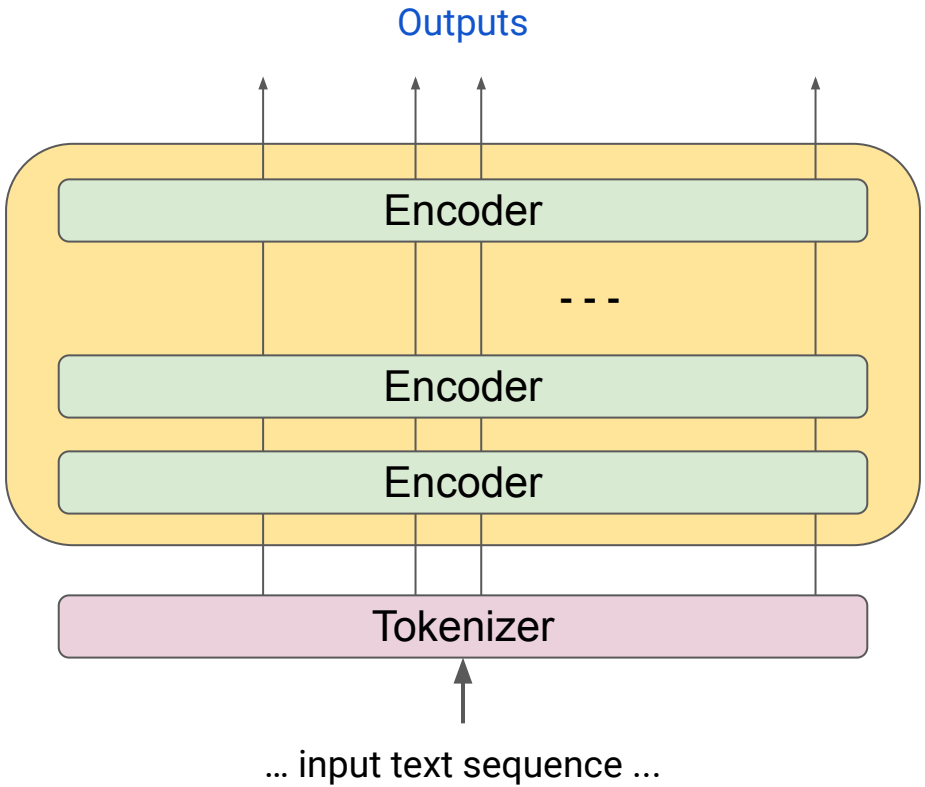


Analytics Vidhya

BERT for Sequence Prediction



BERT Tokenizer



BERT Tokenizer

- Tokenizes the text

BERT Tokenizer

- Tokenizes the text
- Performs contextual encoding and positional encoding

BERT Tokenizer

- Tokenizes the text
- Performs contextual encoding and positional encoding
- Performs padding

BERT Tokenizer

- Tokenizes the text
- Performs contextual encoding and positional encoding
- Performs padding
- Add Special tokens [CLS], [SEP], and [PAD]

Analytics Vidhya

BERT Tokenizer

- Tokenizes the text
- Performs contextual encoding and positional encoding
- Performs padding
- Add Special tokens [CLS], [SEP], and [PAD]
- Convert tokenized sequences to integer sequences

Analytics Vidhya

How to Use BERT?

- Train the model from scratch

How to Use BERT?

- Train the model from scratch
  - Requires huge amount of text data
  - Resource intensive model training

How to Use BERT?

- Train the model from scratch
    - Requires huge amount of text data
    - Resource intensive model training
- Fine-tune a pre-trained model

How to Use BERT?

- Train the model from scratch
  - Requires huge amount of text data
  - Resource intensive model training
- Fine-tune a pre-trained model
  - Works well on smaller datasets
  - Can be trained on modest systems such as Google Colab

Analytics Vidhya

Thank You

Analytics
Vidhya