

Language Models

A *Language Model* is an instance of the "predict the next" paradigm where

- given a sequence of words
- we try to predict the next word

Recall the architecture to solve "predict the next word" and data preparation

Language Modeling task

Architecture

Data preparation

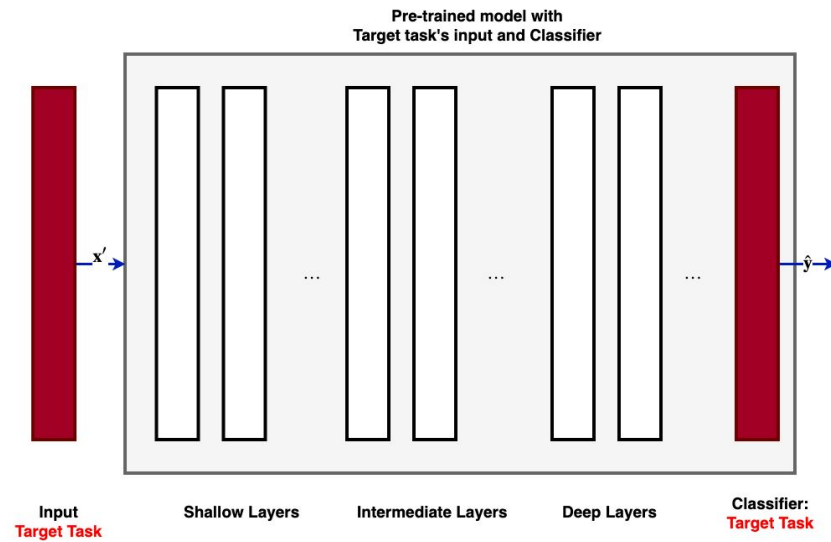
Fine-Tuning

Logically, we use the process that we described as Transfer Learning

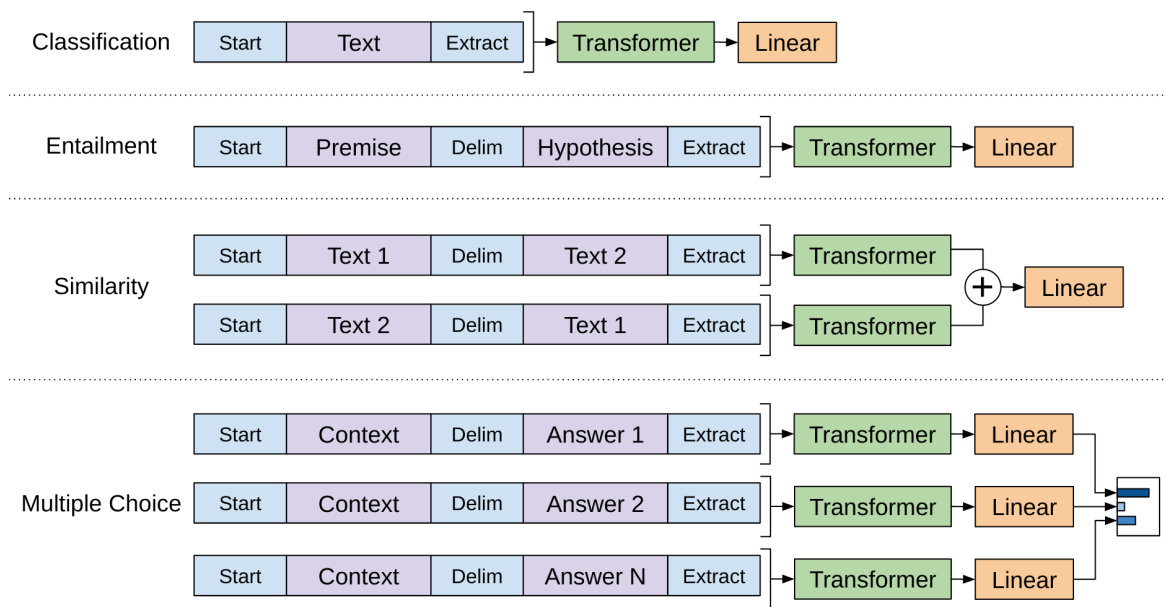
- where we use the output of some layer of the Pre-Trained model
 - default: all layers, excluding the Classification Head
- as a "meaningful" **fixed length** representation of input sequence $\mathbf{x}_{(1)}^{(i)}, \dots, \mathbf{x}_{(m)}^{(i)}$
- which is then fed to a Classification head with the object of matching the target $\mathbf{y}^{(i)}$

Recall the diagram from our module on [Transfer Learning \(Transfer_Learning.ipynb\)](#)

Transfer Learning: replace the head of the pre-trained model



GPT: Task encoding



Picture from: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

• Similarity

- Input is a *pair* (or more) of text sequences [First, Second, Third]
- Binary/Multinomial classification: Probability that other sentences similar to First?

First: Machine Learning is easy not hard
 Second: Machine Learning is not difficult
 Third: Machine Learning is hard not easy
 Label: [Second: .95, Third: .01]

To use the Pre-Trained LM + Fine-Tuning approach

- we need to convert the structured input into simple sequences.

See [this paper \(https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf\)](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf) for some transformations.

