Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 75%. We keep your highest score.

Next item

1. In transformer architectures, excluding the translation task, what is the principal distinction between encoders and decoders? 1 / 1 point

- ○ Encoders apply a SoftMax function before generating the output.
- ◉ Decoders leverage an attention mechanism involving matrix multiplication.
- ○ Encoders generate text by predicting the previous tokens in a sequence.
- ○ Decoders include a multi-head attention mechanism, but encoders do not.

Correct
The principal distinction between encoders and decoders in transformer architectures lies in the use of masked self-attention for decoders. Decoders leverage an attention mechanism, which at its core involves matrix multiplication.

2. Which of the following statements is true for the characteristics of a decoder model implemented using PyTorch with a causal language model (LM) architecture? 1 / 1 point

- ○ The decoder models in PyTorch using causal LM restrict attending sequential data due to its limitations in capturing temporary tokens.
- ○ The decoder models in PyTorch using causal LM initially focus on understanding the context from bidirectional dependencies within a sequence.
- ◉ The decoder models in PyTorch using causal LM help to generate sequence-to-sequence tokens while attending to the previous tokens and ensuring coherence in sequence generation.
- ○ The decoder models in PyTorch use causal LM relays on input embeddings without considering the order of tokens in the sequence.

Correct
A decoder model in PyTorch with causal LM only attends the previously generated tokens during sequence-to-sequence generation tokens.

3. Which of the following best describes the purpose of using decoder model neural network architecture? 1 / 1 point

- ◉ Generate outputs from the learned representations of the input data.
- ○ Initialize the weights of the neural network parameters.
- ○ Preprocess the input data before feeding it to the neural network.
- ○ Reduce the dimensionality of the data for faster computation.

Correct
Decoder models are useful to generate output from the learned representations of the input data, such as image generation, language translation, and sequence generation.

4. Which of the following is one of the advantages of implementing causal attention masking in natural language processing (NLP)? 1 / 1 point

- ◉ Language translation
- ○ Sentiment analysis
- ○ Speech recognition
- ○ Image classification

Correct
Language translation helps in the translation process to maintain the chronological order of the word to maintain the meaning of the sentence.