Congratulations! You passed!

Grade received 80%
Latest Submission Grade 80%
To pass 80% or higher

Go to next item

1. A Transformer Network, unlike its predecessors RNNs, GRUs and LSTMs, can process entire sentences all at the same time. (Parallel architecture). 1 / 1 point

- (•) True
- ( ) False

Expand

Correct
A Transformer Network can ingest entire sentences all at the same time.

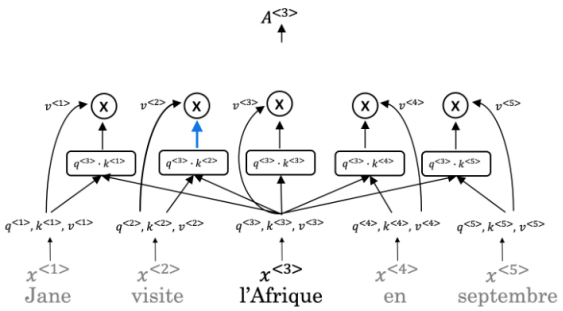
2. Transformer Network methodology is taken from: 1 / 1 point

- (•) Attention Mechanism and CNN style of processing.
- ( ) Attention Mechanism and RNN style of processing.
- ( ) RNN and LSTMs
- ( ) GRUs and LSTMs

Expand

Correct
Transformer architecture combines the use of attention based representations and a CNN convolutional neural network style of processing.

3. How does the Self-Attention mechanism of transformers use neighboring words to compute a word's context? 1 / 1 point



- ( ) Selecting the maximum word values to map the Attention related to that given word.
- ( ) Multiplication of the word values to map the Attention related to that given word.
- ( ) Selecting the minimum word values to map the Attention related to that given word.
- (•) Summation of the word values to map the Attention related to that given word.

Expand

Correct
Given a word, its neighboring words are used to compute its context by summing up the word values to map the Attention related to that given word.

4. What letter does the "?" represent in the following representation of *Attention*? 1 / 1 point

$$Attention(Q, K, V) = softmax(\frac{QK^?}{\sqrt{d_?}})V$$

- ( ) q
- (•) k

- t
- v

Expand

Correct
k is represented by the ? in the representation.

5. Which of the following statements represents Key (K) as used in the self-attention calculation? **0 / 1 point**

- K = qualities of words given a Q
- K = the order of the words in a sentence
- K = interesting questions about the words in a sentence
- K = specific representations of words given a Q

Expand

Incorrect
To revise the concept watch the lecture ; V = specific representations of words given a Q

6. $\boldsymbol{Attention(W_i^Q Q, W_i^K K, W_i^V V)}$ **1 / 1 point**

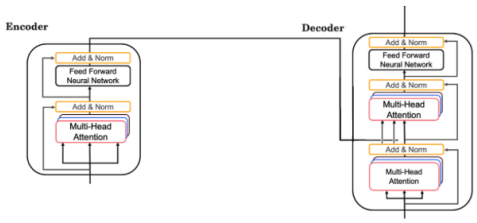
i here represents the computed attention weight matrix associated with the ith "word" in a sentence.

- False
- True

Expand

Correct
Correct! i here represents the computed attention weight matrix associated with the ith "head" (sequence).

7. Following is the architecture within a Transformer Network ***(without displaying positional encoding and output layers(s)).*** **1 / 1 point**



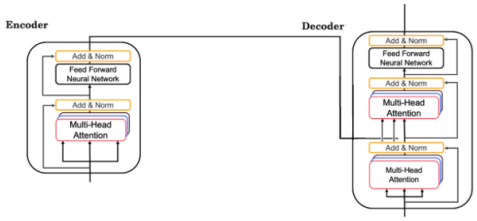
What is ***NOT*** necessary for the *Decoder's* second block of *Multi-Head Attention*?

- K
- All of the above are necessary for the Decoder's second block.
- Q
- V

Expand

Correct
The first block's output is used to generate the Q matrix for the next Multi-Head Attention block. The Decoder also uses K and V from the Encoder for its second block of Multi-Head Attention.

8. Following is the architecture within a Transformer Network ***(without displaying positional encoding and output layers(s)).*** **1 / 1 point**



The output of the decoder block contains a softmax layer followed by a linear layer to predict the next word one word at a time.

- ○ True
- ◉ False

Expand

Correct
The output of the decoder block contains a linear layer followed by a softmax layer to predict the next word one word at a time.

9. Why is positional encoding important in the translation process? (Check all that apply) **1 / 1 point**

- ☑ Position and word order are essential in sentence construction of any language.

Correct

- ☐ It helps to locate every word within a sentence.
- ☐ It is used in CNN and works well there.
- ☑ Providing extra information to our model.

Correct

Expand

Correct
Great, you got all the right answers.

10. Which of these is a good criterion for a good positional encoding algorithm? **0 / 1 point**

- ☐ It should output a unique encoding for each time-step (word's position in a sentence).
- ☐ Distance between any two time-steps should be consistent for all sentence lengths.
- ☐ The algorithm should be able to generalize to longer sentences.
- ☑ None of these.

This should not be selected

Expand

Incorrect
You didn't select all the correct answers