Congratulations! You passed!

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

Go to next item

1. A Transformer Network processes sentences from left to right, one word at a time. **1 / 1 point**

- ○ True
- ◉ False

Expand

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

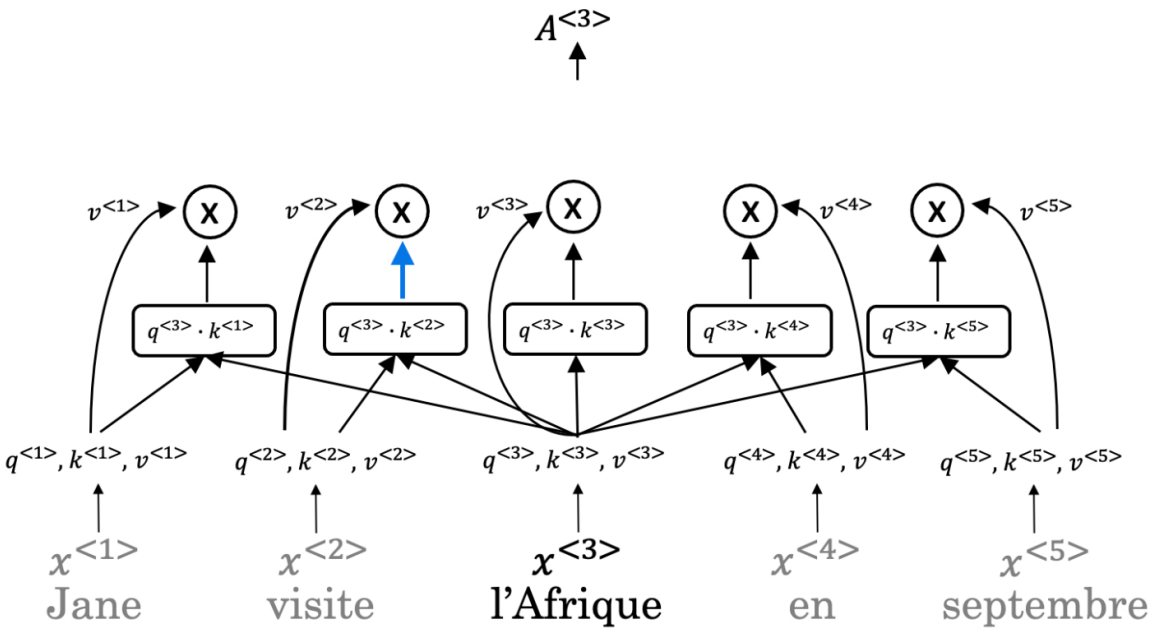
2. Transformer Network methodology is taken from: **1 / 1 point**

- ◉ Attention Mechanism and CNN style of processing.
- ○ RNN and LSTMs
- ○ Attention Mechanism and RNN style of processing.
- ○ 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.
- Summation 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.

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. Which of the following correctly represents *Attention*? **0 / 1 point**

- $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q * k^{i})} {{\sum}_j\exp(q * k^{j}))* V^{i}$$
- $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q * v^{i})} {{\sum}_j\exp(q * v^{j}))* K^{i}$$
- $${A(Q,K,V)} = (\frac{\exp(q * k^{i})} {\exp(q * k^{j}))* V^{i}$$
- $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q * k^{i})} {{\sum}_j\exp(q * k^{j}))* {\sum}_i{v}^{i}$$

Expand

Incorrect
To review the Attention formula watch the lecture *Self-Attention*.

5. Are the following statements true regarding Query (Q), Key (K) and Value (V)? **1 / 1 point**

Q = interesting questions about the words in a sentence

K = qualities of words given a Q

V = specific representations of words given a Q

- (•) True
- ( ) False

Expand

Correct
Q = interesting questions about the words in a sentence, K = qualities of words given a Q, V = specific representations of words given a Q

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

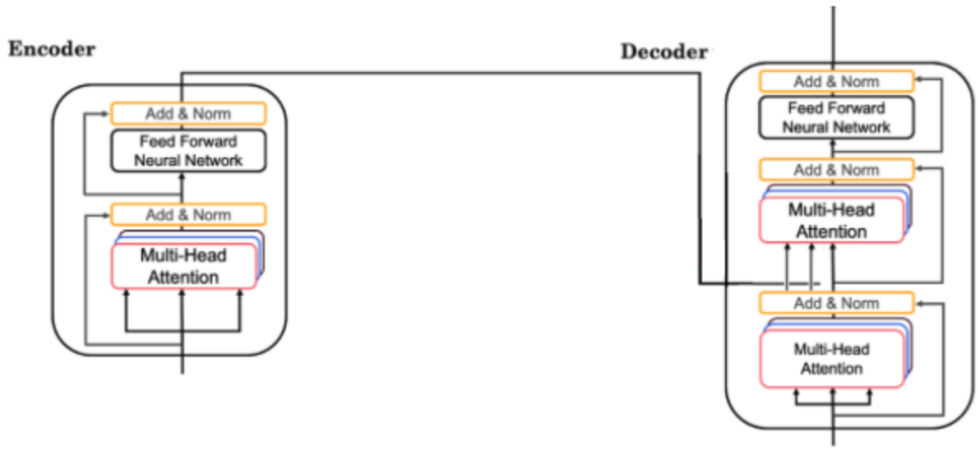
i here represents the computed attention weight matrix associated with the ith "head" (sequence).

- ( ) False
- (•) True

Expand

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)).*** **0 / 1 point**



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

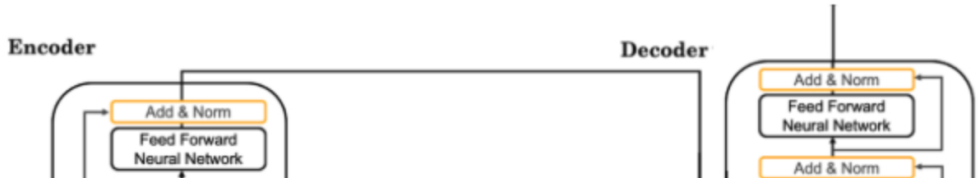
- V
- Q
- All of the above are necessary for the Decoder's second block.
- K

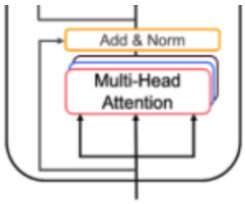
Expand

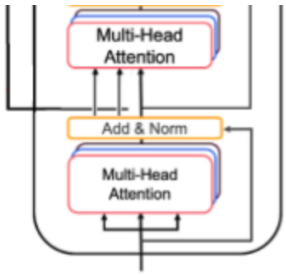
Incorrect

The first block's output is used to generate the Q matrix for the next Multi-Head Attention block. To revise the concept watch the lecture *Transformer Network*.

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.

- (●) False
- ( ) True

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. Which of the following statements is true about positional encoding? Select all that apply. **1 / 1 point**

- [x] Positional encoding provides extra information to our model.

Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

- [ ] Positional encoding is used in the transformer network and the attention model.
- [x] Positional encoding is important because position and word order are essential in sentence construction of any language.

Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

- [x] Positional encoding uses a combination of sine and cosine equations.

Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

Expand

Correct
Great, you got all the right answers.

10. Which of these is ***not*** a good criterion for a good positional encoding algorithm? **1 / 1 point**

- (●) It should output a common encoding for each time-step (word's position in a sentence).
- ( ) Distance between any two time-steps should be consistent for all sentence lengths.
- ( ) It must be deterministic.
- ( ) The algorithm should be able to generalize to longer sentences.

Expand

Correct
This is not a good criterion for a good positional encoding algorithm.