Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

1. A Transformer Network, like its predecessors RNNs, GRUs and LSTMs, can process information one word at a time. (Sequential architecture). 1 / 1 point

- ○ True
- ◉ False

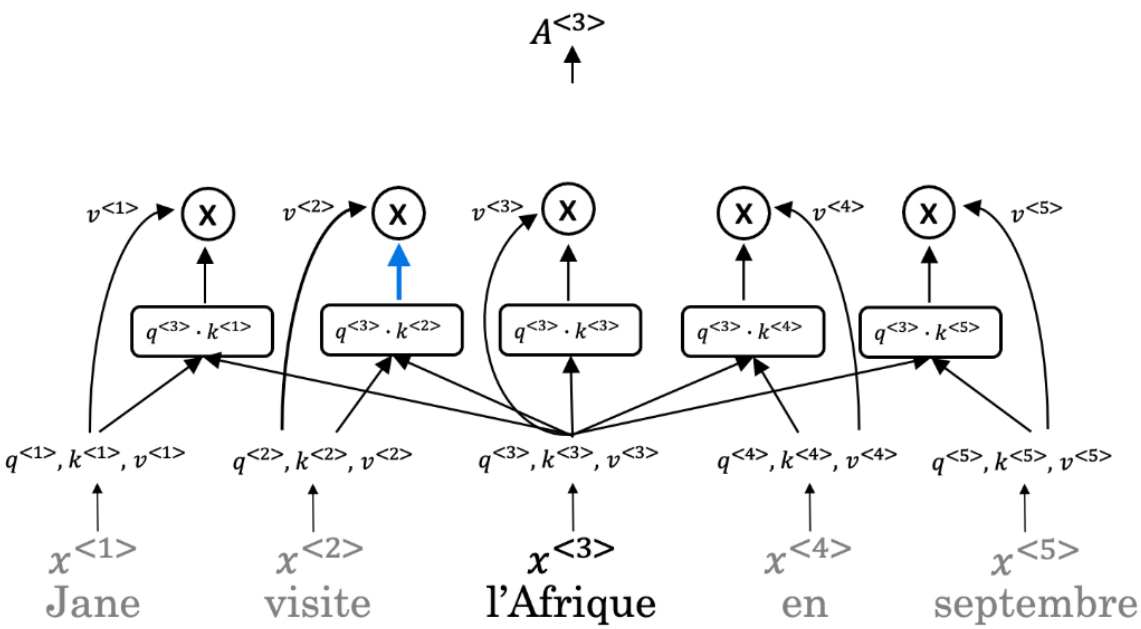
Correct
Correct! A Transformer Network can ingest entire sentences all at the same time.

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

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

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

3. What are the key inputs to computing the attention value for each word? 1 / 1 point



- ○ The key inputs to computing the attention value for each word are called the quotation, knowledge, and value.
- ○ The key inputs to computing the attention value for each word are called the query, knowledge, and vector.
- ◉ The key inputs to computing the attention value for each word are called the query, key, and value.
- ○ The key inputs to computing the attention value for each word are called the quotation, key, and vector.

Correct
The key inputs to computing the attention value for each word are called the query, key, and value.

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

- $Attention(Q, K, V) = min(\frac{QK^T}{\sqrt{d_k}})V$
- $Attention(Q, K, V) = min(\frac{QV^T}{\sqrt{d_k}})K$
- $Attention(Q, K, V) = softmax(\frac{QV^T}{\sqrt{d_k}})K$
- (selected) $Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$

Correct

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 = specific representations of words given a Q

V = qualities of words given a Q

- (selected) False
- True

Correct
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

$$\boldsymbol{Attention(W_i^Q Q, W_i^K K, W_i^V V)}$$

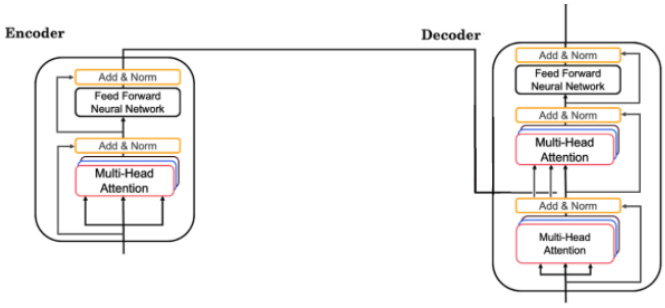
1 / 1 point

6. What does i represent in this multi-head attention computation?

- The computed attention weight matrix associated with the order of the words in a sentence
- The computed attention weight matrix associated with specific representations of words given a Q
- (selected) The computed attention weight matrix associated with the *ith* "head" (sequence)
- The computed attention weight matrix associated with the *ith* "word" in a sentence.

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

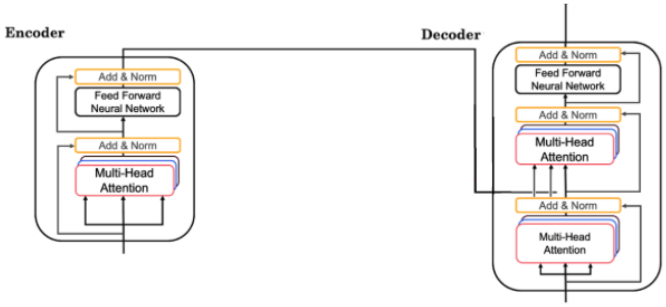


What is ***NOT*** necessary for the *Decoder's* second block of *Multi-Head Attention*?

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

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

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

- ☑ 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 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.*

- ☐ Positional encoding is used in the transformer network and the attention model.
- ☑ 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.*

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

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

Correct
This is a good criterion for a good positional encoding algorithm.