Your grade: 90%

Your latest: **90%** • Your highest: **90%** • 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 (selected)

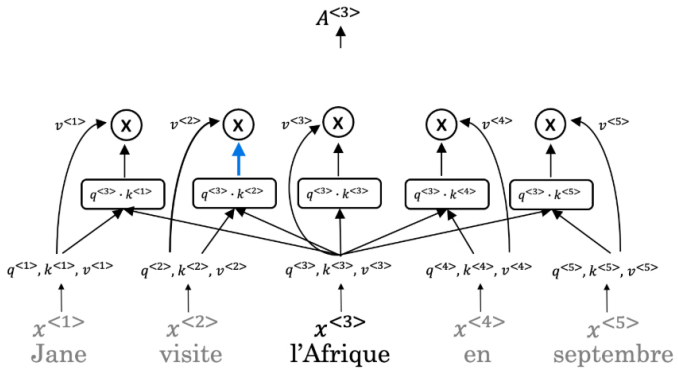
Correct
Correct! A Transformer Network can ingest entire sentences all at the same time.

2. The major innovation of the transformer architecture is combining the use of LSTMs and RNN sequential processing. 1 / 1 point

- True
- False (selected)

Correct
The major innovation of the transformer architecture is combining the use of attention based representations and a CNN convolutional neural network style of processing.

3. The concept of *Self-Attention* is that: 1 / 1 point



- Given a word, its neighbouring words are used to compute its context by taking the average of those word values to map the Attention related to that given word.
- Given a word, its neighbouring words are used to compute its context by selecting the lowest of those word values to map the Attention related to that given word.
- Given a word, its neighbouring words are used to compute its context by selecting the highest of those word values to map the Attention related to that given word.
- Given a word, its neighbouring words are used to compute its context by summing up the word values to map the Attention related to that given word. (selected)

Correct

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

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

- v
- k (selected)
- q
- t

Correct
k is represented by the ? in the representation.

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

- (●) 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

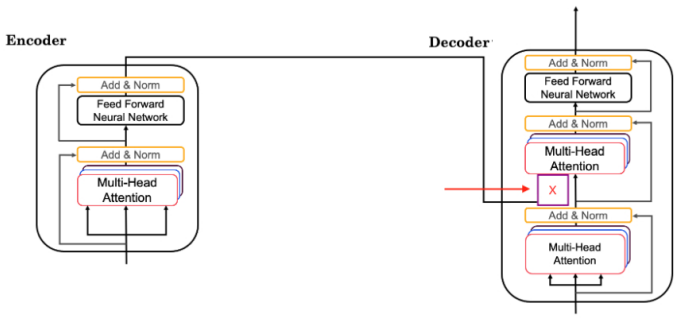
6. i here represents the computed attention weight matrix associated with the ith "word" in a sentence.

- ( ) True
- (●) False

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 information does the *Decoder*take from the *Encoder* for its second block of *Multi-HeadAttention*? (Marked X, pointed by the independent arrow)

(Check all that apply)

- [ ] Q
- [x] K

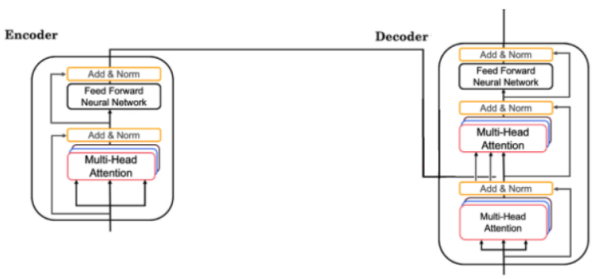
Correct

- [x] V

Correct

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

1 / 1 point



What does the output of the *encoder* block contain?

- ○ Softmax layer followed by a linear layer.
- ◉ Contextual semantic embedding and positional encoding information
- ○ Prediction of the next word.
- ○ Linear layer followed by a softmax layer.

Correct
The output of the *encoder* block contains contextual semantic embedding and positional encoding information.

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

10. Which of these is ***not*** a good criterion for a good positional encoding algorithm? **0 / 1 point**

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

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