Congratulations! You passed!

Grade received 100% Latest Submission Grade 100% To pass 80% or higher

Go to 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

- (•) False
- ( ) True

Expand

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

2. Transformer Network methodology is taken from: (Check all that apply) 1 / 1 point

- [ ] Convolutional Neural Network style of architecture.
- [x] Attention mechanism.

Correct

- [x] Convolutional Neural Network style of processing.

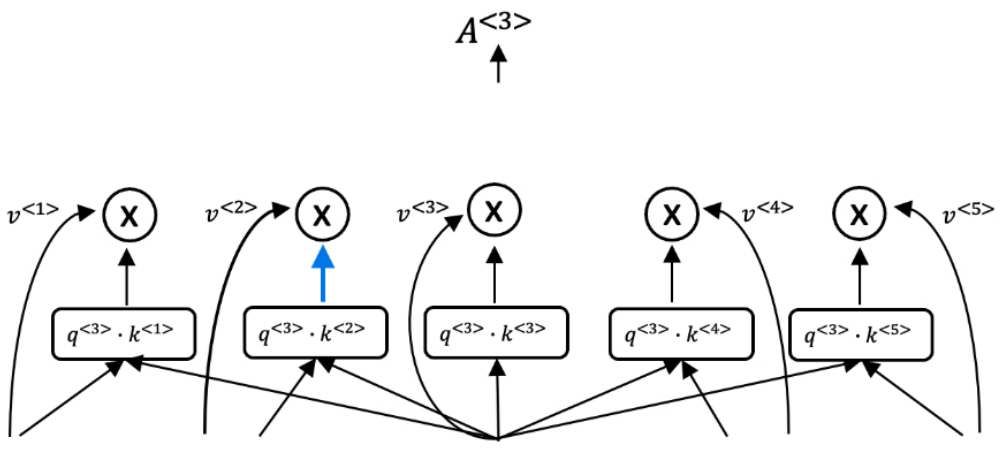
Correct

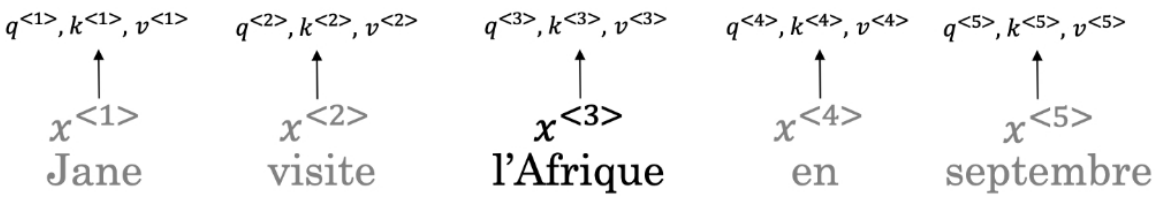
- [ ] None of these.

Expand

Correct
Great, you got all the right answers.

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





- 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 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 summing up the 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.

Expand

Correct

4. Which of the following correctly represents *Attention ?*

1 / 1 point

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

Expand

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 = qualities of words given a Q

V = specific representations of words given a Q

- False
- True

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

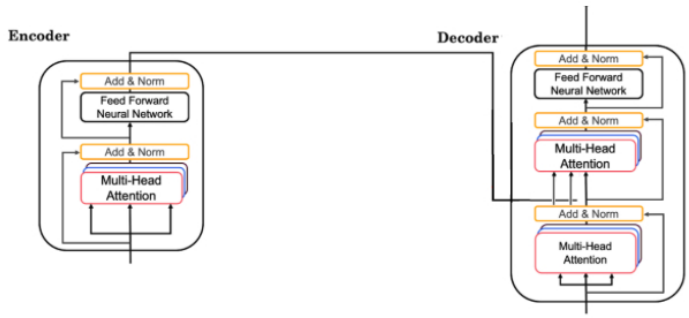
i here represents the computed attention weight matrix associated with the ith "head" (sequence).

- 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)).*** 1 / 1 point



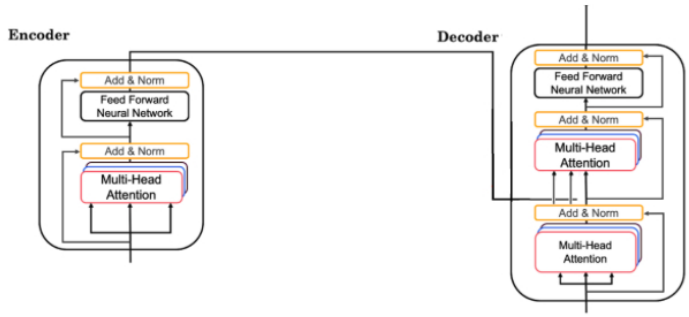
What is generated from the output of the *Decoder's* first block of *Multi-Head Attention*?

- Q
- V
- K

Expand

Correct
This first block's output is used to generate the Q matrix for the next Multi-Head Attention block.

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

- [x] 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.
- [x] 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? **1 / 1 point**

- [x] It should output a unique encoding for each time-step (word's position in a sentence).

✓ **Correct**

- [x] Distance between any two time-steps should be consistent for all sentence lengths.

✓ **Correct**

- [x] The algorithm should be able to generalize to longer sentences.

✓ **Correct**

- [ ] None of these.

Expand

Correct
Great, you got all the right answers.