Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **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**

- (•) False
- ( ) True

Expand

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**

- [x] Convolutional Neural Network style of processing.

Correct

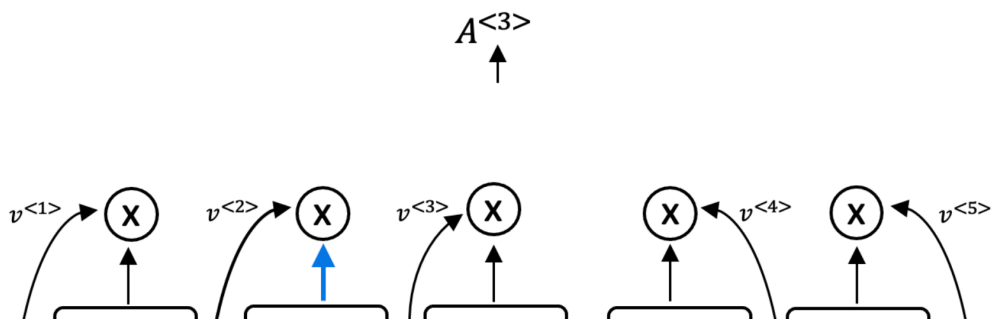
- [ ] None of these.
- [ ] Convolutional Neural Network style of architecture.
- [x] Attention mechanism.

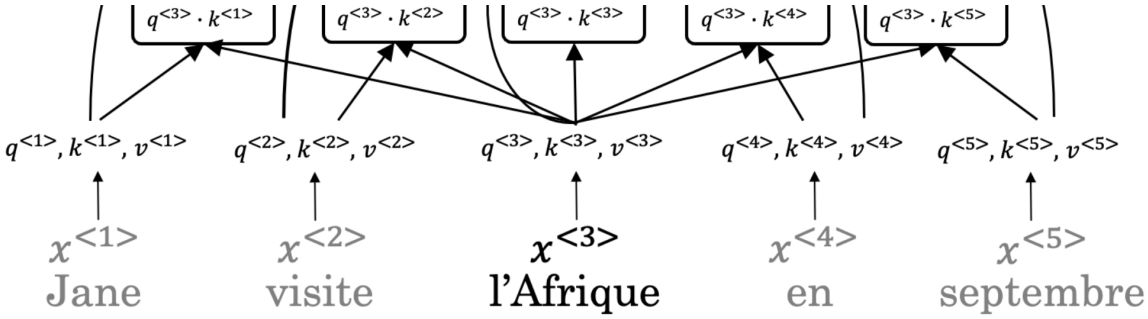
Correct

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 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.
- 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.

Expand

Correct

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

1 / 1 point

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

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

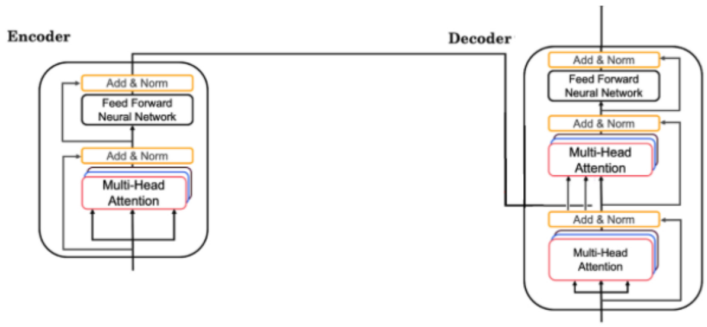
i here represents the computed attention weight matrix associated with the ith "word" in a sentence.

- True
- False

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



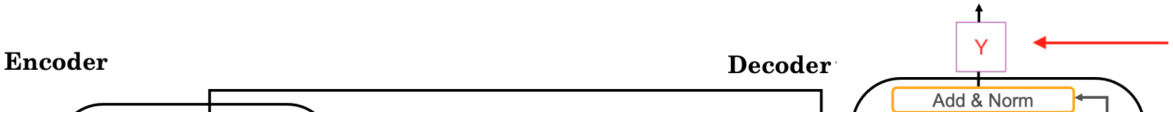
What is generated from the output of the *Decoder's* first block of *Multi-Head Attention*?

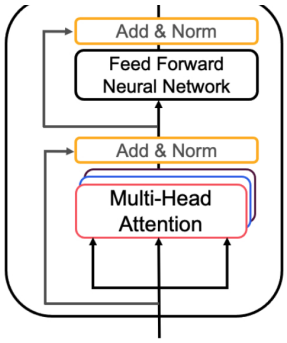
- K
- V
- Q

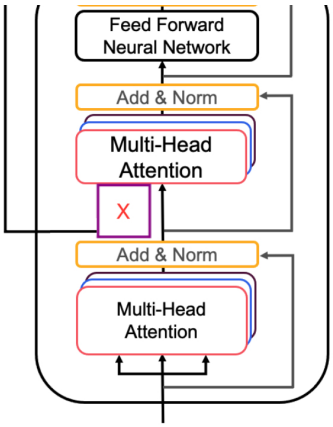
Expand

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







What is the output layer(s) of the *Decoder*? (Marked Y, pointed by the independent arrow)

- ○ Softmax layer
- ◉ Linear layer followed by a softmax layer.
- ○ Linear layer
- ○ Softmax layer followed by a linear layer.

Expand

✓ Correct

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 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*.

- ☑ 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.

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**

- ☑ 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.