Congratulations! You passed!

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

Retake the assignment in **23h 57m**

Go to next item

1. A Transformer Network, unlike its predecessors RNNs, GRUs and LSTMs, can process entire sentences all at the same time. (Parallel architecture). **1 / 1 point**

- False
- True

Expand

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

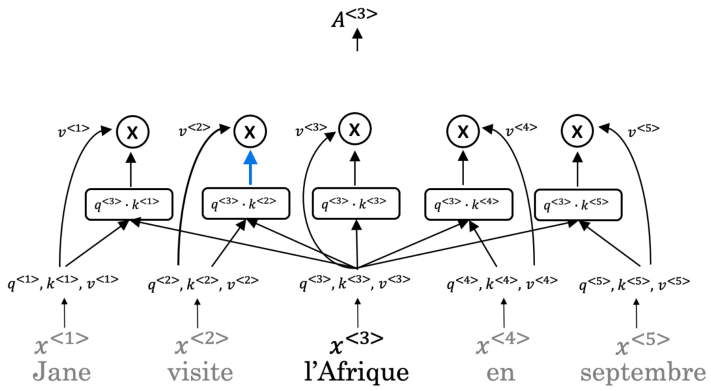
2. Transformer Network methodology is taken from: **1 / 1 point**

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

Expand

Correct
Transformer architecture combines 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 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**

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

Expand

Correct
This is the correct Attention formula.

5. Which of the following statements represents Key (K) as used in the self-attention calculation? **1 / 1 point**

- K = specific representations of words given a Q
- K = interesting questions about the words in a sentence
- K = the order of the words in a sentence
- K = qualities of words given a Q

Expand

Correct
The qualities of words given a Q are represented by Key (K).

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

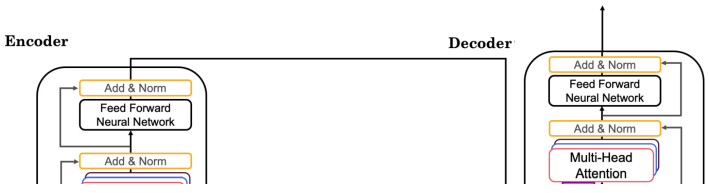
What does i represent in this multi-head attention computation?

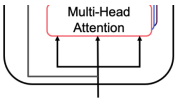
- The computed attention weight matrix associated with the order of the words in a sentence
- The computed attention weight matrix associated with specific representations of words given a Q
- The computed attention weight matrix associated with the *ith* "word" in a sentence.
- The computed attention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

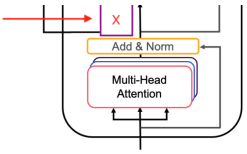
Expand

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

(Check all that apply)

- [ ] Q
- [x] V

Correct

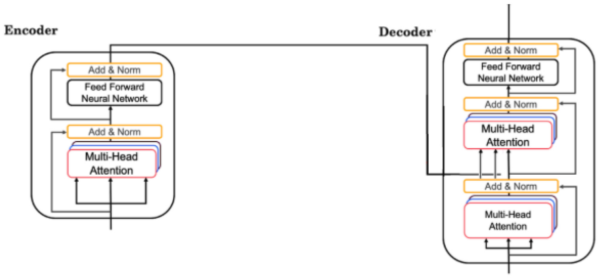
- [x] K

Correct

Expand

Correct
Great, you got all the right answers.

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

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*.

- [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*.

- [ ] 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 ***not*** a good criterion for a good positional encoding algorithm? **1 / 1 point**

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

Expand

Correct

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