

Sequence Models

Week 4: Transformers

- 1 A Transformer Network processes sentences from left to right, one word at a time.

Answer: False.

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

- 2 Transformer Network methodology is taken from:

Answer: Attention Mechanism and CNN style of processing.

Comment: 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?

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

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

- 4 What letter does the k represent in the following representation of Attention? $Attention(Q, K, V) = softmax(Qk^T / \sqrt{d_k})V$.

Answer: k.

- 5 Are the following statements true regarding Query (Q), Key (K) and Value (V)? Q = interesting questions about the words in a sentence; K = specific representations of words given a Q; V = qualities of words given a Q.

Answer: False.

Comment: 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 $Attention(W_i^Q Q, W_i^K K, W_i^V V)$ i here represents the computed attention weight matrix associated with the ith word in a sentence.

Answer: False.

Comment: i here represents the computed attention weight matrix associated with the ith head sequence.

- 7 Following is the architecture within a Transformer Network. What information does the Decoder take from the Encoder for its second block of Multi Head Attention? (Check all that apply)

Answer: K; V.

- 8 Following is the architecture within a Transformer Network. (without displaying positional encoding and output layers(s)) What is the output layer(s) of the Decoder? (Marked Y pointed by the independent arrow)

Answer: Linear layer followed by a softmax layer.

9 Why is positional encoding important in the translation process? (Check all that apply)

Answer: Position and word order are essential in sentence construction of any language;
Providing extras information to our model.

10 Which of these is a good criteria for a good positional encoding algorithm?

Answer: It should output a unique encoding for each time step (word's position in a sentence); Distance between any two time steps should be consistent for all sentence lengths;
The algorithm should be able to generalize to longer sentences.