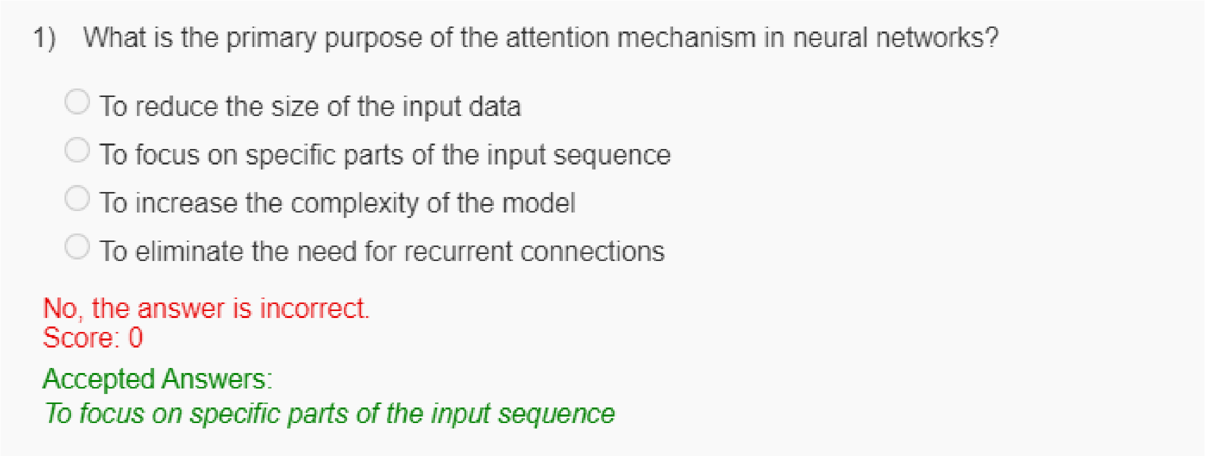


1) What is the primary purpose of the attention mechanism in neural networks?

○ To reduce the size of the input data
○ To focus on specific parts of the input sequence
○ To increase the complexity of the model
○ To eliminate the need for recurrent connections

No, the answer is incorrect.
Score: 0

Accepted Answers:
To focus on specific parts of the input sequence

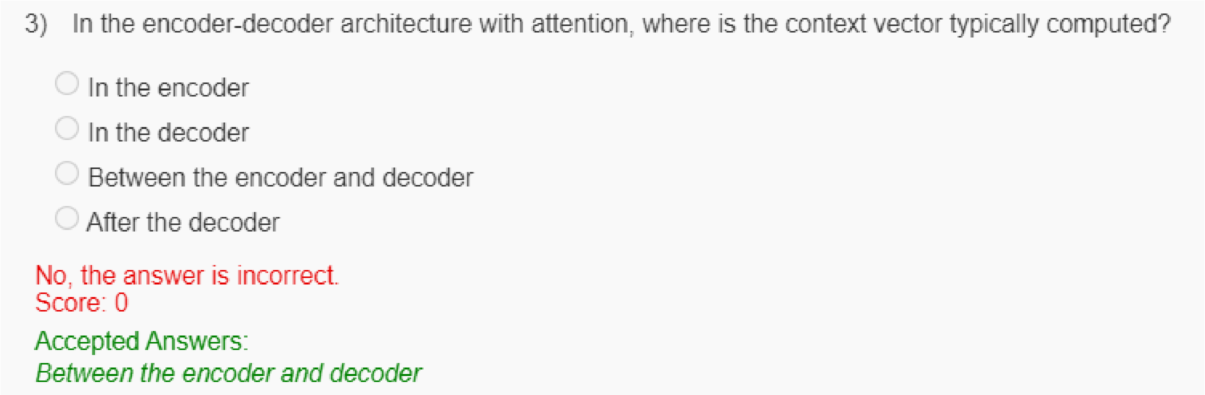
The scenarios that would most benefit from hierarchical attention mechanisms are **summarizing long text documents**. Hierarchical attention mechanisms are designed to handle data with multiple levels of structure, such as documents that contain paragraphs, sentences, and words, allowing the model to focus on relevant parts at different levels of abstraction.

Hierarchical attention mechanisms are best for summarizing long text documents because they handle multi-level structures (e.g., paragraphs, sentences).

For the other scenarios:

1. **Classifying images**: Uses CNNs, not hierarchical attention.
2. **Analyzing customer reviews**: Usually doesn't require hierarchical structure; simpler attention may suffice.
3. **Real-time processing of sensor data**: Tends to focus on sequential rather than hierarchical features.

Thus, hierarchical attention is most beneficial for text-oriented tasks that have layered information.



3) In the encoder-decoder architecture with attention, where is the context vector typically computed?

○ In the encoder
○ In the decoder
○ Between the encoder and decoder
○ After the decoder

No, the answer is incorrect.
Score: 0

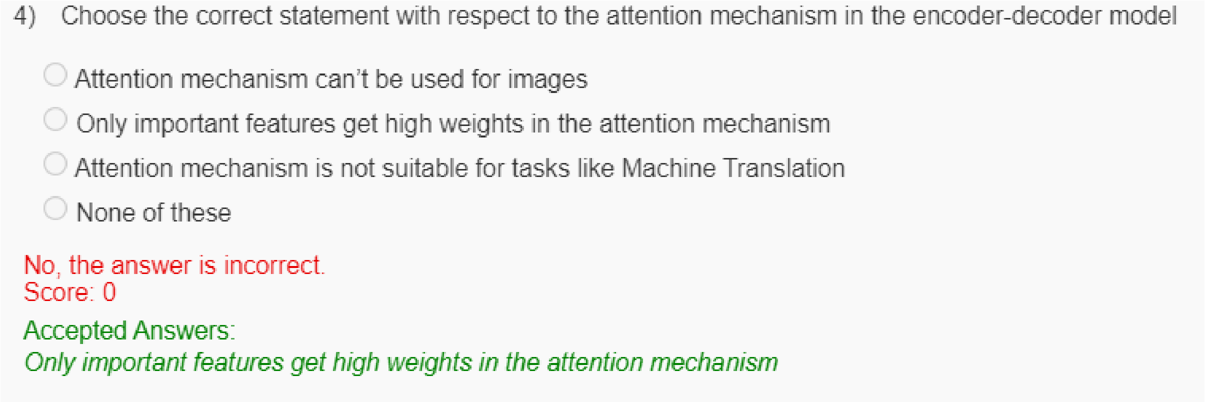
Accepted Answers:
Between the encoder and decoder

Correct Option: Between the encoder and decoder

Why not other options:

1. **In the encoder**: The encoder generates hidden states from the input sequence but does not compute the context vector. The context vector is formed based on these hidden states via the attention mechanism.
2. **In the decoder**: The decoder uses the context vector to generate the output sequence, but it does not compute it. The computation is based on the attention mechanism that references the encoder's outputs in conjunction with the decoder's current state.
3. **After the decoder**: The context vector must be computed during the decoding process to provide relevant information for generating the output. Computing it after the decoding would not allow the decoder to utilize this information effectively.

In conclusion, the context vector is computed **between the encoder and decoder**, utilizing the outputs from the encoder based on the attention mechanism.



4) Choose the correct statement with respect to the attention mechanism in the encoder-decoder model

○ Attention mechanism can’t be used for images
○ Only important features get high weights in the attention mechanism
○ Attention mechanism is not suitable for tasks like Machine Translation
○ None of these

No, the answer is incorrect.
Score: 0

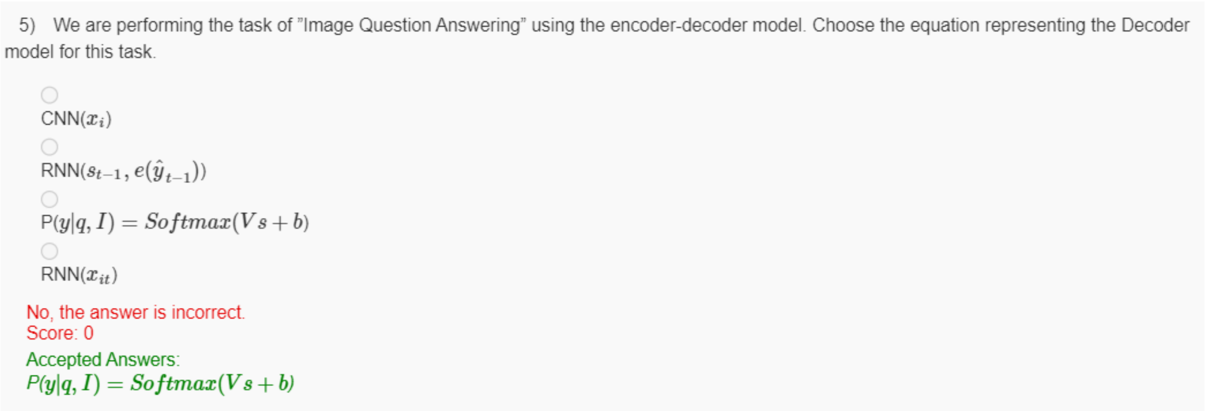
Accepted Answers:
Only important features get high weights in the attention mechanism

The correct statement with respect to the attention mechanism in the encoder-decoder model is: **Only important features get high weights in the attention mechanism.**

Why the other options are incorrect:

1. **Attention mechanism can’t be used for images**: This is false. Attention mechanisms can indeed be used for images, such as in image captioning and object detection tasks, where attention helps focus on relevant parts of the image.
2. **Attention mechanism is not suitable for tasks like Machine Translation**: This is also false. In fact, attention mechanisms are particularly well-suited for tasks like machine translation, as they help the model focus on relevant parts of the input sentence during the decoding process.
3. **None of these**: Since one of the statements is correct ("Only important features get high weights in the attention mechanism"), this choice is also incorrect.

Thus, the correct answer is that **only important features get high weights in the attention mechanism**.

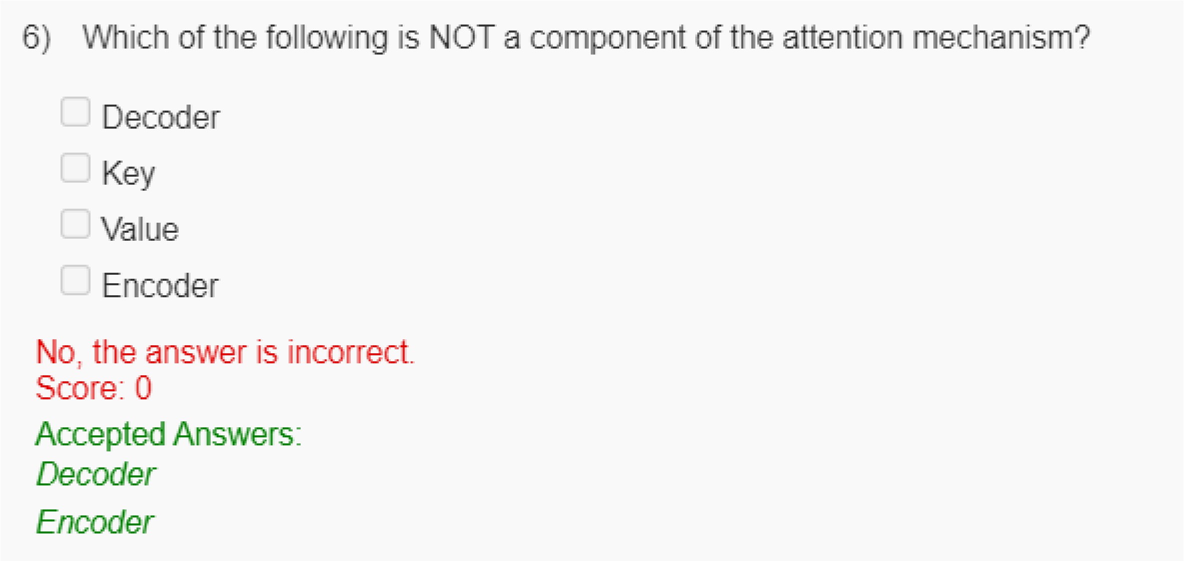


5) We are performing the task of "Image Question Answering" using the encoder-decoder model. Choose the equation representing the Decoder model for this task.

○ CNN(x_i)
○ RNN($s_{t-1}, e(\hat{y}_{t-1})$)
○ P($y|q, I) = Softmax(Vs + b)$
○ RNN(x_{it})

No, the answer is incorrect.
Score: 0

Accepted Answers:
P($y|q, I) = Softmax(Vs + b)$

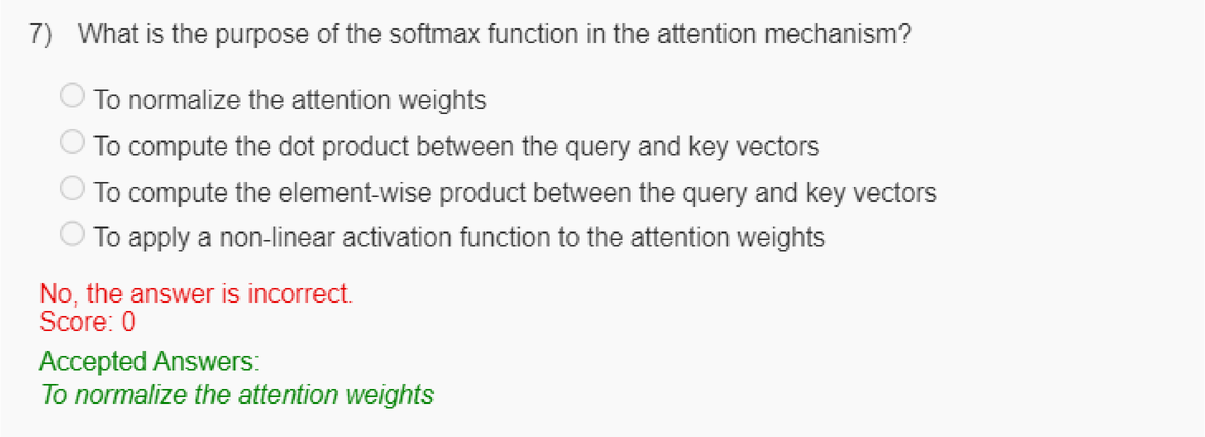


6) Which of the following is NOT a component of the attention mechanism?

☐ Decoder
☐ Key
☐ Value
☐ Encoder

No, the answer is incorrect.
Score: 0

Accepted Answers:
Decoder
Encoder



7) What is the purpose of the softmax function in the attention mechanism?

○ To normalize the attention weights
○ To compute the dot product between the query and key vectors
○ To compute the element-wise product between the query and key vectors
○ To apply a non-linear activation function to the attention weights

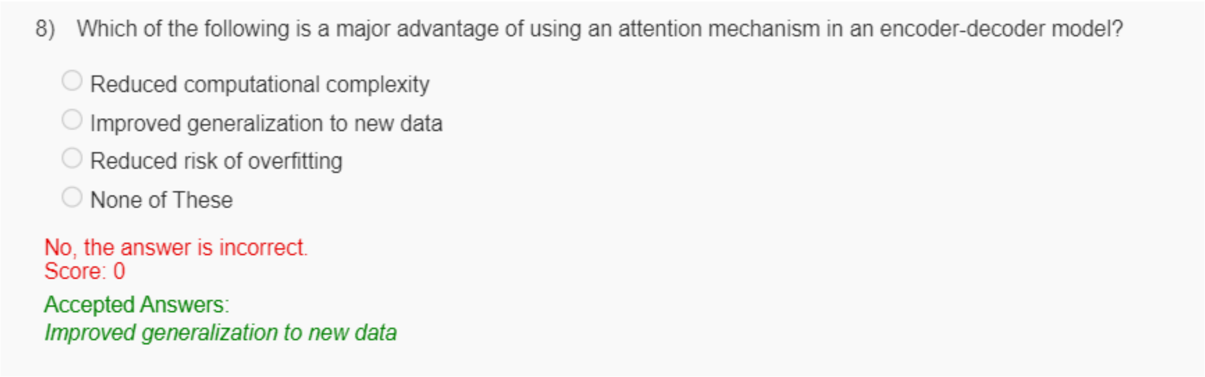
No, the answer is incorrect.
Score: 0

Accepted Answers:
To normalize the attention weights

Apologies for any confusion in my previous response. To clarify, the attention mechanism specifically includes the components of **Key**, **Value**, and **Query**. The **Encoder** and **Decoder** parts of the larger encoder-decoder architecture, but they themselves are not components the attention mechanism.

So, the correct answer to the question **"Which of the following is NOT a component of the attention mechanism?"** can indeed include both **Encoder** and **Decoder**.

1. **To normalize the attention weights**: This is the correct answer. The softmax function transforms the raw attention scores (which can be any real numbers) into a probability distribution. This means all the attention weights will sum up to one, allowing them to be interpreted as probabilities that reflect how much attention should be paid to each input element.
2. **To compute the dot product between the query and key vectors**: This is not correct. The dot product between the query and key vectors is computed to get raw attention scores before applying the softmax function, not the purpose of the softmax itself.
3. **To compute the element-wise product between the query and key vectors**: This is incorrect. The softmax function does not compute the element-wise product; rather, it is applied after obtaining the attention scores to normalize those scores.
4. **To apply a non-linear activation function to the attention weights**: This statement is misleading. While the softmax function can be viewed as a transformation, its primary function is normalization, not simply applying a non-linear activation in the same way as other activation functions (like ReLU or sigmoid).

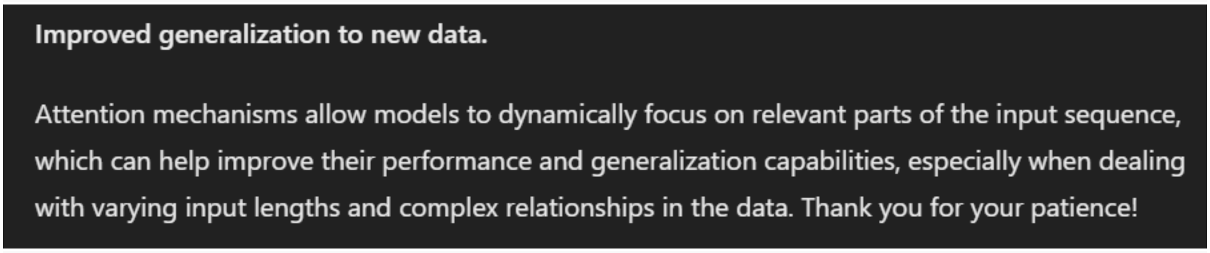


8) Which of the following is a major advantage of using an attention mechanism in an encoder-decoder model?

○ Reduced computational complexity
○ Improved generalization to new data
○ Reduced risk of overfitting
○ None of These

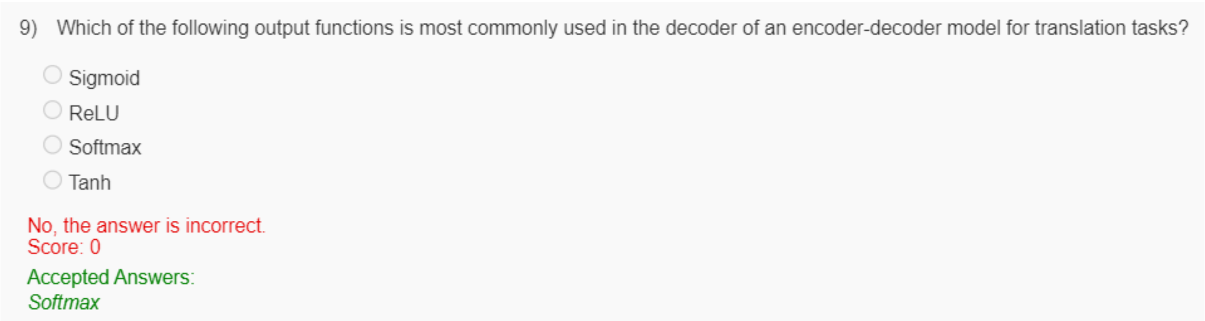
No, the answer is incorrect.
Score: 0

Accepted Answers:
Improved generalization to new data



Improved generalization to new data.

Attention mechanisms allow models to dynamically focus on relevant parts of the input sequence, which can help improve their performance and generalization capabilities, especially when dealing with varying input lengths and complex relationships in the data. Thank you for your patience!

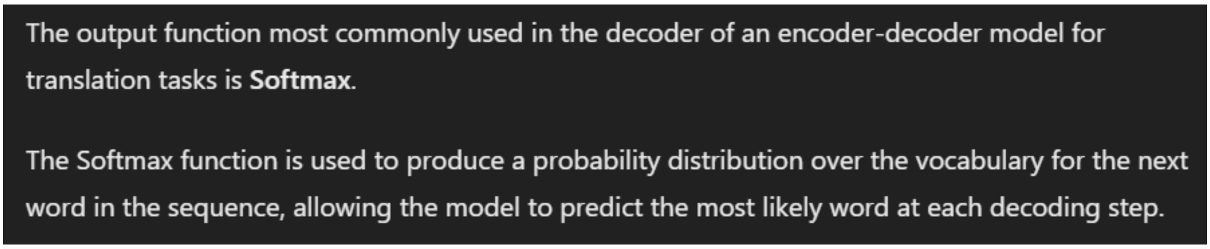


9) Which of the following output functions is most commonly used in the decoder of an encoder-decoder model for translation tasks?

○ Sigmoid
○ ReLU
○ Softmax
○ Tanh

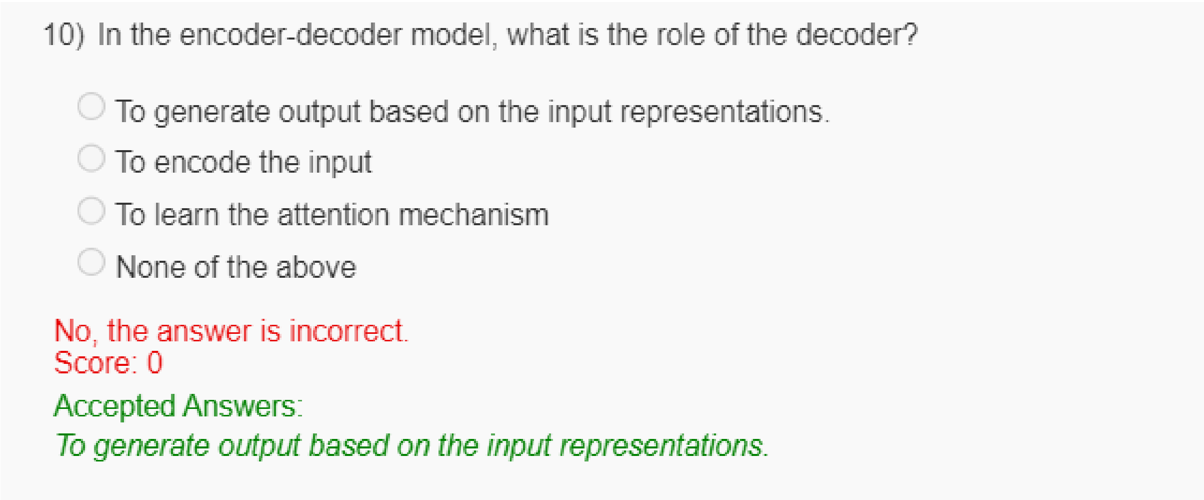
No, the answer is incorrect.
Score: 0

Accepted Answers:
Softmax



The output function most commonly used in the decoder of an encoder-decoder model for translation tasks is **Softmax**.

The Softmax function is used to produce a probability distribution over the vocabulary for the next word in the sequence, allowing the model to predict the most likely word at each decoding step.



10) In the encoder-decoder model, what is the role of the decoder?

○ To generate output based on the input representations.
○ To encode the input
○ To learn the attention mechanism
○ None of the above

No, the answer is incorrect.
Score: 0

Accepted Answers:
To generate output based on the input representations.