## Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 60%. We keep your highest score.

Next item

1. In a transformer model's self-attention mechanism, consider a token at position 't' is 'apple' in a sentence 'I love to eat apples every day.' You've applied a self-attention mechanism to predict the token 'apple' representation based on the context provided in the sentence. Which of the following is the most influential factor for the token 'apple' prediction in the self-attention mechanism? 1 / 1 point

- ○ The word 'apple'
- ○ The word 'love'
- ○ The presence of the word 'eat'
- ◉ Position of the token 'apple' in the sequence

**Correct**
In the self-attention mechanism, the token's position within the sentence plays an important role in determining the context effectively.

2. In the following embedding, what is the dimensionality of the embedding for the word Transformers? 1 / 1 point


- ○ 1
- ◉ 4
- ○ 0.2
- ○ 3

**Correct**
All the words in this sequence have a dimensionality of 4.

3. What purpose does the following formula serve in the context of using the attention mechanism in language translation? 1 / 1 point

$$\hat{w} = \text{argmax}_i \{\mathbf{h}\mathbf{V}^T\}$$

- ○ Finds the index of the embedding that is more similar to H.
- ◉ Retrieves the translated word from the translated vector
- ○ Provides the word vector for the translated word
- ○ Applies attention mechanism to word embeddings

**Correct**
To retrieve the translated word from the translated vector, you can apply the given formula, where h is the output of the attention mechanism and V is the value matrix. You get a vector of zeros except for the one that is the same. This column has the same index as the word index of the translated word in the value matrix.

4. Which statement is true about the scaled dot-product attention mechanism with multiple heads? 1 / 1 point

- ○ There is a constraint that the input dimension must be a prime number.
- ◉ Each head can attend to distinct segments of the input sequence in parallel.
- ○ It follows the setting batch_first = True for PyTorch implementations.
- ○ It is implemented using the nn.Attention module of PyTorch.

**Correct**
Multi-head attention operates by executing several scaled dot-product attention processes in parallel. This strategy allows each head to attend to distinct segments of the input sequence.

5. When using transformer-based models for text classification, creating the text pipeline is a key activity. Identify the missing step (step number 5) from the following list of steps for creating the text pipeline. 1 / 1 point

Steps for creating the text pipeline:

1. Create iterators and allocate a training set
2. Generate tokens and construct a vocabulary
3. Design a custom collate function
4. Apply padding
5. ?

- ○ Add positional encoding
- ○ Record cumulative losses and epoch accuracies
- ◉ Create a data loader
- ○ Apply the transformer encoder layers

**Correct**
As the next step in the text pipeline creation process, you will create a data loader for the training, validation, and testing sets.