|  |  |
| --- | --- |
| **Started on** | Thursday, April 1, 2021, 2:16 PM |
| **State** | Finished |
| **Completed on** | Thursday, April 1, 2021, 2:42 PM |
| **Time taken** | 26 mins 20 secs |
| **Points** | 7.00/7.00 |
| **Grade** | **100.00** out of 100.00 |

Top of Form

Question **1**

Correct

1.00 points out of 1.00

Flag question

Question text

Which of the following is a benefit of transformers that vanilla RNNs and LSTMs lacked?

Select one:

a. The ability to be processed in parallel

b. Created a memory

c. Did not treat data as independent

d. Allowed for forgetting and remembering different tokens

Feedback

Your answer is correct.

The correct answer is: The ability to be processed in parallel

Question **2**

Correct

1.00 points out of 1.00

Flag question

Question text

What is different about the attention you apply to the inputs of the encoder versus the attention you apply to the inputs of the decoder?

Select one:

a. The encoder model's attention output is not normalized

b. The attention on the decoder input is masked

c. You don't apply attention to the inputs of the decoder

d. The attention in the encoder is multiheaded, the attention in the decoder is not

Feedback

Your answer is correct.

The correct answer is: The attention on the decoder input is masked

Question **3**

Correct

1.00 points out of 1.00

Flag question

Question text

In traditional word2vec word embeddings, you have one embedding for every word, regardless of how many meanings that word has in practice. Which innovation in NLP made it possible to distinguish between different meanings for the same word?

Select one:

a. Contextual word embeddings

b. Softmax

c. Linear layer

d. Transformers

Feedback

Your answer is correct.

The correct answer is: Contextual word embeddings

Question **4**

Correct

1.00 points out of 1.00

Flag question

Question text

In practice, which kind of processing do most NLP practitioners use when doing gradient descent?

Select one:

a. NLP practitioners do not use gradient descent for their machine learning models

b. Stochastic gradient descent

c. Full gradient descent

d. Gradient descent with batching

Feedback

Your answer is correct.

The correct answer is: Gradient descent with batching

Question **5**

Correct

1.00 points out of 1.00

Flag question

Question text

Which of the following is a solution to having inputs of different lengths in your batch calculations?

Select one:

a. Conditional Random Field

b. There is no problem with having inputs of different lengths in your batch, that only matters for gradient descent

c. Padding

d. Masking

Feedback

Your answer is correct.

The correct answer is: Padding

Question **6**

Correct

1.00 points out of 1.00

Flag question

Question text

True or False: if you use padding, you don't need to keep track of your original input lengths.

Select one:

a. False

b. True

Feedback

Your answer is correct.

The correct answer is: False

Question **7**

Correct

1.00 points out of 1.00

Flag question

Question text

Which of the following are things you already know about that would be helpful in many NLP contexts?

Select one:

a. Support Vector Machines

b. Clustering

c. All of the above

d. Cosine similarity

Feedback

Your answer is correct.

The correct answer is: All of the above