Quiz for NLP-preprocessing-tensorflow

Tuesday, June 25, 2019 6:39 PM

1
point

1. Question 1

What is the name of the TensorFlow library containing common data that you can use to train and test neural networks?

- [ ] TensorFlow Data Libraries
- [ ] There is no library of common data sets, you have to use your own
- [x] TensorFlow Datasets
- [ ] TensorFlow Data

Question 2

1
point

2. Question 2

How many reviews are there in the IMDB dataset and how are they split?

- [x] 50,000 records, 50/50 train/test split
- [ ] 50,000 records, 80/20 train/test split
- [ ] 60,000 records, 50/50 train/test split
- [ ] 60,000 records, 80/20 train/test split

Question 3

1
point

3. Question 3

How are the labels for the IMDB dataset encoded?

- [ ] Reviews encoded as a number 1-5
- [x] Reviews encoded as a number 0-1
- [ ] Reviews encoded as a boolean true/false
- [ ] Reviews encoded as a number 1-10

Question 4

1

point

4. Question 4

What is the purpose of the embedding dimension?

- [x] It is the number of dimensions for the vector representing the word encoding
- [ ] It is the number of words to encode in the embedding
- [ ] It is the number of dimensions required to encode every word in the corpus
- [ ] It is the number of letters in the word, denoting the size of the encoding

Question 5

1
point

5. Question 5

When tokenizing a corpus, what does the num_words=n parameter do?

- [ ] It specifies the maximum number of words to be tokenized, and picks the first 'n' words that were tokenized
- [ ] It errors out if there are more than n distinct words in the corpus
- [x] It specifies the maximum number of words to be tokenized, and picks the most common 'n' words
- [ ] It specifies the maximum number of words to be tokenized, and stops tokenizing when it reaches n

Question 6

1
point

6. Question 6

To use word embeddings in TensorFlow, in a sequential layer, what is the name of the class?

- [ ] tf.keras.layers.Embed
- [x] tf.keras.layers.Embedding
- [ ] tf.keras.layers.WordEmbedding
- [ ] tf.keras.layers.Word2Vector

Question 7

1
point

7. Question 7

IMDB Reviews are either positive or negative. What type of loss function should be used in this scenario?

- [ ] Categorical crossentropy
- [ ] Binary crossentropy

- [x] Binary crossentropy
- [ ] Adam
- [ ] Binary Gradient descent

Question 8

1
point

8. Question 8

When using IMDB Sub Words dataset, our results in classification were poor. Why?

- [ ] The sub words make no sense, so can't be classified
- [x] Sequence becomes much more important when dealing with subwords, but we're ignoring word positions
- [ ] We didn't train long enough
- [ ] Our neural network didn't have enough layers