0/22 Questions Answered

Final Exam 1399

Q1 Training/Validation Loss
4 Points

If the training loss is much lower than the validation loss while training a DNN classifier, what could be the problem?

Enter your answer here

Which one of the following could address the issue?

Reduce the number of layers

Decrease Dropout Probability

Increase L2 regularization weight

Increase the size of each hidden layer

Save Answer

Q2 Zero Shot Learning
6 Points

Explain what is zero shot learning.

Enter your answer here

Give an example.

Enter your answer here

Save Answer

Q3 Recursive/Recurrent Neural Networks
3 Points

Explain the difference between Recurrent and Recursive Neural Networks.

Enter your answer here

Recursive NN example:

Enter your answer here

Recurrent NN example:

Enter your answer here

Save Answer

Q4 Multi-Head attention in BERT
5 Points

Using BERT Explorer tool (https://huggingface.co/exbert/) choose the bert-base-cased model. Provide a sample input sentence (must write-in your own sentence) and a task and explain why having more than one attention head is useful.

Explanation:

Enter your answer here

Screenshot(s):

Please select file(s) Select file(s)

Save Answer

Q5 BERT Layers
12 Points

Q5.1 English
6 Points

If you were to use BERT embeddings for classification, how would you choose the layer?

Enter your answer here

Using BERT Explorer tool (https://huggingface.co/exbert/) choose the bert-base-cased model. Provide a sample input sentence (must write-in your own sentence) and a task and explain why one of the layers between 1-6 would do better. Provide explanation as well as screenshot(s) of the tool.

Explanation:

Enter your answer here

Screenshot(s):

Please select file(s) Select file(s)

Now provide a sample input sentence and a task and explain why one of the layers between 6 and 12 would be better for embeddings in that task.

Explanation:

Enter your answer here

Screenshot(s):

Please select file(s) Select file(s)

Save Answer

Q5.2 Farsi
6 Points

Repeat the previous question with the distilbert-base-multilingual-cased model and a Farsi input sentence (must write-in your own sentence).
Layers 1- 3:

Explanation:

Enter your answer here

Screenshot(s):

Please select file(s) Select file(s)

Layers 4-6:
Explanation:

Enter your answer here

Screenshot(s):

Please select file(s) Select file(s)

Save Answer

Q6 Embeddings
23 Points

Q6.1 One-Hot Vector
3 Points

When using one-hot encoding (or just using a unique number) for words, is the training and testing affected by the choice of number for each word?

Yes

No

Explain why? Provide a concrete example(with actual embeddings). Make assumptions as needed.

Enter your answer here

Save Answer

Q6.2 Word2Vec
6 Points

Let's say we have a sentiment analysis task with "user reviews" as input and positive, negative and neutral classes as output. Would using Word2Vec embeddings instead of one-hot encoding provide any benefit?

Yes

No

Explain why? Provide a concrete example(with actual embeddings). Make assumptions as needed.

Enter your answer here

What data should Word2Vec be trained on for optimal performance?

Enter your answer here

Save Answer

Q6.3 BERT
8 Points

For the same task as previous question, would using BERT embeddings instead of Word2Vec provide any benefit?

Yes

No

Explain why? Provide a concrete example(with actual embeddings). Make assumptions as needed.

Enter your answer here

What is the main difference between Word2Vec and BERT embeddings?

Enter your answer here

What data should BERT be trained on for optimal performance?

Enter your answer here

Explain what the purpose of the CLS token in BERT.

Enter your answer here

Save Answer

Q6.4 Character Embeddings like FastText
6 Points

For the same task as previous question, would using character embeddings like FastText instead of BERT provide any benefit?

Yes

No

Does it have any disadvantages?

Yes

No

Explain why? Provide a concrete example (with actual embeddings). Make assumptions as needed.

Enter your answer here

Save Answer

Q7 Language Models
10 Points

Q7.1 Usage
4 Points

Language Models can be used for two main tasks. What are they?

Enter your answer here

Save Answer

Q7.2 Intrinsic Evaluation
3 Points

Explain what Intrinsic Evaluation is and how language models can be intrinsically evaluated. If you mention a specific metric, provide details about how its value can be interpreted.

Enter your answer here

Save Answer

Q7.3 Extrinsic Evaluation
3 Points

Explain what Extrinsic Evaluation is and how language models can be extrinsically evaluated. If you mention a specific metric, provide details about how its value can be interpreted.

Enter your answer here

Save Answer

Q8 Feature Engineering
21 Points

Suppose you are developing a Named Entity Recognizer for medical domain. You notice many named entities in this domain have specific word/character shapes. For example Varicella-zoster, mRNA, CPA1. Based on this observation you decide to add a custom word-shape feature which captures this information explicitly. You replace all uppercase characters with X, lowercase with x and digits with d and remove all repetition while always keeping the first and last character. For example, Varicella-zoster becomes Xx-x, mRNA becoms xX and CPA1 becomes Xd.

Q8.1 Would it help?

3 Points

Explain why this additional feature could help improve the named entity recognizer's accuracy.

Enter your answer here

Explain why (and in what circumstances) it may not be needed.

Enter your answer here

Explain why it may be a bad idea to add this feature.

Enter your answer here

Save Answer

Q8.2 Baseline Model
12 Points

Suppose we have decided to find out for ourselves whether this custom feature is helpful or not. Propose a baseline architecture that is capable of performing named entity recognition. At training time, the model takes in a sentence. Each word is marked as Inside or Outside. If it is marked as Inside it also has a category like PERSON, LOCATION, DRUG. At test time an input sentence is provided and the Inside/Outside labels and category is expected.
Provide a schematic of your model along with an explanation.

Model Schematic:

Please select file(s) Select file(s)

Explanation:

Enter your answer here

Save Answer

Q8.3 Feature Encoding
6 Points

Explain how you would incorporate the additional word-shape feature into your model. Provide exact and specific details about input/output vectors.

Enter your answer here

Provide an example of how the input vectors would change:

Enter your answer here

Save Answer

Q9 Softmax
14 Points

If you don't remember Softmax details, you can visit here:
https://en.wikipedia.org/wiki/Softmax_function

Q9.1 Code
2 Points

Write a function in python to compute the softmax vector given logits in an input list, logits. You cannot use the numpy library for this. Upload the python file.

Please select file(s) Select file(s)

Save Answer

Q9.2 Values
2 Points

Enter your student ID here:

Enter your answer here

Let's say the logits for the output layer of your neural network are the last 4 digits of your student id. Run your code above and enter the softmax values below (in order and separated by space).

Enter your answer here

Save Answer

Q9.3 Temperature
3 Points

Update your code above to include a Temperature parameter. Upload the python file.

Please select file(s) Select file(s)

Save Answer

Q9.4 Values with temperature
3 Points

Provide the softmax value from the logits above for various temperatures:
T = 1

Enter your answer here

T = 10

Enter your answer here

T = 100

Enter your answer here

Save Answer

Q9.5 Purpose
2 Points

Explain in your own word, what is temperature in softmax and how it could/should be used.

Enter your answer here

Save Answer

Q9.6 Student-Teacher Compression
2 Points

Explain how the "Temperature" is used in Student-Teacher model compression. You don't have to be exact. The rough idea is enough.

Enter your answer here

Save Answer

Save All Answers

Submit & View Submission ❯