

## 复查测验提交: Homework 3

用户	生物医学工程学院 吉泓光
课程	自然语言处理
测试	Homework 3
已开始	24-4-14 上午5:37
已提交	24-4-15 下午10:45
截止日期	24-4-18 下午11:59
状态	已完成
尝试分数	得 136.66667 分, 满分 140 分
已用时间	41 小时 7 分钟
显示的结果	所有答案, 已提交的答案, 正确答案

### 问题 1

得 10 分, 满分 10 分

BERT is finetuned using MLM (masked language modeling) and NSP (next sentence prediction ) and then pretrained on downstream tasks such as text classification.

所选答案: ☒ 错

答案: ☐ 对

☒ 错

### 问题 2

得 10 分, 满分 10 分

Unlike BERT, ELMO is **NOT** a kind of contextualized word embedding.

所选答案: ☒ 错

答案: ☐ 对

☒ 错

### 问题 3

得 10 分, 满分 10 分

ELMO is based on Bidirectional LSTM. Therefore, we cannot use ELMO to finish language modeling tasks.

所选答案: ☒ 错

答案: 对  
☒ 错

#### 问题 4

得 10 分, 满分 10 分

BERT is a kind of non-autoregressive model.

所选答案: ☒ 对  
答案: ☒ 对  
☐ 错

#### 问题 5

得 6.66667 分, 满分 10 分

Please select all the correct statement(s)

所选 ☒ A.  
答 Parameter Efficient Tuning only updates a small number of parameters, which  
案: reduces memory cost.  
☒ D.  
For LoRA, it will not lead to additional inference latency because LoRA components  
can be added to original weight matrices after tuning.  
答 ☒ A.  
案: Parameter Efficient Tuning only updates a small number of parameters, which  
reduces memory cost.  
☒ B.  
Prefix Tuning does not prepend the trainable parameters to the input text as a prefix  
directly, so it is different from Prompt Tuning.  
C.  
For Adaptor, additional Adapter Network will be added only before Multi-Head  
Attention.  
☒ D.  
For LoRA, it will not lead to additional inference latency because LoRA components  
can be added to original weight matrices after tuning.

#### 问题 6

得 10 分, 满分 10 分

Masked language modeling task is one of the tasks used to pre-train GPT

所选答案: ☒ 错  
答案: 对  
☒ 错

#### 问题 7

得 10 分, 满分 10 分

Select the correct state for Prompt

所选答 ☒ A.

案： Prompt Tuning can help us utilize a pre-trained language model in downstream tasks more efficiently than fine-tuning.

答案： ☒ A.  
Prompt Tuning can help us utilize a pre-trained language model in downstream tasks more efficiently than fine-tuning.

B. Prompt is a kind of tunable parameter.

C. Prompt Tuning is a task used in the pre-training procedure of GPT models.

D. We need to update the parameters of the whole model to finish prompt tuning.

## 问题 8

得 10 分，满分 10 分

Please select all the correct statement(s).

所选 ☒ A.

答案： Transformer-based LLMs can roughly be divided into three categories: encoder-only models, decoder-only models and encoder-decoder models.

☒ C.

Retrieval Augmented Generation (RAG) demonstrates potential in interpretability and knowledge updating.

答案： ☒ A.

Transformer-based LLMs can roughly be divided into three categories: encoder-only models, decoder-only models and encoder-decoder models.

B.

LLMs without finetuning are good at text generation, especially when up-to-date factual knowledge is needed.

☒ C.

Retrieval Augmented Generation (RAG) demonstrates potential in interpretability and knowledge updating.

D.

Nowadays, training corpora for LLM are still restricted to plain text, not including code.

## 问题 9

得 10 分，满分 10 分

Select all the correct statement(s)

所选答

☒ B. There is a cross-attention module in encoder-decoder language models.

案：

☒ D.

About 1.5% of input words are left unchanged in BERT's pre-training procedure.

答案：

☐ A. GPT is a kind of bidirectional language model.

☒ B.

There is a cross-attention module in encoder-decoder language models.

☐ C. T5 is a classical decoder-only pre-trained language model like GPT.

☒ D.

About 1.5% of input words are left unchanged in BERT's pre-training procedure.

## 问题 10

得 10 分，满分 10 分

Please select all correct statement(s)

所选 ☒ A. Reinforcement Learning can be used to pre-trained language models.

答案:

所选 ☒ A. Reinforcement Learning can be used to pre-trained language models.

答案:

B.

In-context learning is one of the pre-training tasks for pre-trained language models.

C.

GPT models can handle input sentences of any length, unlike BERT. This is because they are generative language models.

D.

In most situations, we use the [CLS] token's representation to encode spans in a sentence.

## 问题 11

得 10 分, 满分 10 分

Select all correct statements

所选 ☒ 1.

答案: The Baum-Welch algorithm is a special case of the EM algorithm, and can be used for unsupervised learning of HMM parameters.

案:

☒ 3.

The forward-backward algorithm has the same time and space complexity as the Viterbi algorithm.

☒ 4.

We can use the "count and normalize" strategy to train an HMM in both supervised and unsupervised manners. The difference is that, in supervised learning, "count" is the actual counts, whereas in unsupervised learning, "count" is the expected counts.

答案: ☒ 1.

案: The Baum-Welch algorithm is a special case of the EM algorithm, and can be used for unsupervised learning of HMM parameters.

2.

If we use the Baum-Welch algorithm to train an HMM, it will finally converge to a global optimum after running a sufficient number of iterations.

☒ 3.

The forward-backward algorithm has the same time and space complexity as the Viterbi algorithm.

☒ 4.

We can use the "count and normalize" strategy to train an HMM in both supervised and unsupervised manners. The difference is that, in supervised learning, "count" is the actual counts, whereas in unsupervised learning, "count" is the expected counts.

## 问题 12

得 10 分, 满分 10 分

Select all correct statements

所选 ☒

答案: The difference between MEMM and CRF is that the former is locally normalized and suffers from the label bias issue, while the latter is globally normalized.

案:

☒

Neural CRFs use neural networks (e.g., BiLSTMs, Transformers) to compute CRF potential scores.

答案:



The difference between MEMM and CRF is that the former is locally normalized and suffers from the label bias issue, while the latter is globally normalized. MEMMs prefer states with higher number of transitions, thus suffer from the label bias issue.

In CRF training, both max-margin loss and negative log-likelihood loss involve the use of the forward algorithm to compute the partition function.



Neural CRFs use neural networks (e.g., BiLSTMs, Transformers) to compute CRF potential scores.

问题 13

得 10 分, 满分 10 分

Given a HMM with transition (Table 1), emission (Table 2) below. And we assume all hidden states transition to the "STOP" with equal probability. Please select the correct optimal sequence of hidden states for the word sequence "I Love ShanghaiTech":

Table 1: Transition		
$Y_{t-1}$	$P(Y_t Y_{t-1})$	
	$H_0$	$H_1$
START	0.6	0.4
$H_0$	0.2	0.8
$H_1$	0.8	0.2

Table 2: Emission		
$W$	$P(W_j H_i)$	
	$H_0$	$H_1$
I	0.4	0.3
Love	0.3	0.2
ShanghaiTech	0.3	0.5

所选答案: (H<sub>1</sub>, H<sub>0</sub>, H<sub>1</sub>)

答案: (H<sub>1</sub>, H<sub>0</sub>, H<sub>1</sub>)

$(H_1, H_1, H_1)$

$(H_1, H_0, H_0)$

$(H_0, H_0, H_1)$

#### 问题 14

得 10 分, 满分 10 分

Suppose the hidden number in an HMM model is  $N$  and the sentence length is  $L$ . Give the time complexity of the forward-backward algorithm in this model.

所选答案: ☒ A.  $O(N^2L)$

答案: ☒ A.  $O(N^2L)$

B.  $O(NL^2)$

C.  $O(N^2L^2)$

D.  $O(NL)$

2024年5月29日 星期三 下午07时31分54秒 CST

← 确定