Week 7 : Assignment 7

The due date for submitting this assignment has passed.

Due on 2025-03-12, 23:59 IST.

Assignment submitted on 2025-03-09, 17:33 IST

1) Which of the following best describes how ELMo's architecture captures different linguistic properties? ***1 point***

- The model explicitly assigns specific linguistic functions to each layer.
- The lower layers capture syntactic information, while higher layers capture semantic information.
- All layers capture the similar properties.
- ELMo uses a fixed, non-trainable weighting scheme for combining layer-wise representations.

Yes, the answer is correct.
Score: 1
Accepted Answers:
The lower layers capture syntactic information, while higher layers capture semantic information.

2) BERT and BART models differ in their architectures. While BERT is (i)------- model, BART is (ii) -------- one. Select the correct choices for (i) and (ii). ***1 point***

- i: Decoder-only , ii: Encoder-only
- i: Encoder-decoder , ii: Encoder-only
- i: Encoder-only , ii: Encoder-decoder
- i: Decoder-only , ii: Encoder-decoder

Yes, the answer is correct.
Score: 1
Accepted Answers:
i: Encoder-only , ii: Encoder-decoder

3) The pre-training objective for the T5 model is based on: ***1 point***

- Next sentence prediction
- Masked language modelling
- Span corruption and reconstruction

- ○ Predicting the next token

Yes, the answer is correct.
Score: 1

Accepted Answers:
Span corruption and reconstruction

4) Which of the following datasets was used to pretrain the T5 model? ***1 point***

- ○ Wikipedia
- ○ BookCorpus
- ○ Common Crawl
- ◉ C4

Yes, the answer is correct.
Score: 1

Accepted Answers:
C4

5) Which of the following special tokens are introduced in BERT to handle sentence pairs? ***1 point***

- ○ [MASK] and [CLS]
- ◉ [SEP] and [CLS]
- ○ [CLS] and [NEXT]
- ○ [SEP] and [MASK]

Yes, the answer is correct.
Score: 1

Accepted Answers:
[SEP] and [CLS]

6) ELMo and BERT represent two different pre-training strategies for language models. ***2 points***
Which of the following statement(s) about these approaches is/are true?

- ☑ ELMo uses a bi-directional LSTM to pre-train word representations, while BERT uses a transformer encoder with masked language modeling.
- ☐ ELMo provides context-independent word representations, whereas BERT provides context-dependent representations.
- ☐ Pre-training of both ELMo and BERT involve next token prediction.
- ☑ Both ELMo and BERT produce word embeddings that can be fine-tuned for downstream tasks.

Yes, the answer is correct.
Score: 2

Accepted Answers:
ELMo uses a bi-directional LSTM to pre-train word representations, while BERT uses a transformer encoder with masked language modeling.
Both ELMo and BERT produce word embeddings that can be fine-tuned for downstream tasks.

7) Decoder-only models are essentially trained based on probabilistic language modelling. ***1 point***
Which of the following correctly represents the training objective of GPT-style models?

- ○ P(y | x) where x is the input sequence and y is the gold output sequence
- ○ P(x | y) where x is the input sequence and y is the gold output sequence
- ◉ $P(w_t \mid w_{1:t-1})$, where w_t represents the token at position t, and $w_{1:t-1}$ is the sequence of tokens from position 1 to t-1
- ○ $P(w_t \mid w_{1:t+1})$, where wt represents the token at position t, and $w_{1:t+1}$ is the sequence of tokens from position 1 to t+1

Yes, the answer is correct.
Score: 1

Accepted Answers:
$P(w_t \mid w_{1:t-1})$, where w_t represents the token at position t, and $w_{1:t-1}$ is the sequence of tokens from position 1 to t-1

8)

In the previous week, we saw the usage of **einsum** function in numpy as a generalized operation for performing tensor multiplications. Now, consider two matrices: $A = \begin{bmatrix} 1 & 5 \\ 3 & 7 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -1 \\ 4 & 2 \end{bmatrix}$.

Then, what is the output of the following numpy operation?

```
numpy.einsum( 'ij, ij -> ',  A, B)
```

23

Yes, the answer is correct.
Score: 2

Accepted Answers:
(Type: Numeric) 23

2 points