

COMPUTING FOR MEDICINE

QUIZ 2 RUBRICS

1. Which of the following statements is **NOT true** about masked language modelling in BERT?
 - A) It helps the model learn the bidirectional context
 - B) It masks entire sentences during training
 - C) It replaces some tokens with [MASK] for prediction
 - D) It allows understanding of relationships between words

Answer: B) It masks entire sentences during training

2. In the PEGASUS model, the _____ task is used for pre-training, which involves masking the entire _____ to encourage the model to learn abstractive capabilities.

Answer: Gap Sentence Generation (GSG), sentences

3. What is/are distinguishing feature(s) of BERT compared to traditional transformers?

- A) Unidirectional context processing
- B) Masked language modelling (MLM)
- C) Sequence generation from scratch
- D) Use of convolutional layers

Answer: B) Masked language modelling (MLM)

- 4.. Which statement best describes multi-head attention?

- A) It uses multiple heads to optimize learning rate.
- B) It divides attention into different subspaces for diverse representation learning.
- C) It applies attention only at the final layer of the model.
- D) It solely focuses on positional embeddings.

Answer: B) It divides attention into different subspaces for diverse representation learning.

5. _____ is designed to enhance natural language understanding by training on bidirectional context.

Answer: BERT

6. In transformers, what is/are the purpose(s) of applying dropout?

- A) To adjust the learning rate dynamically
- B) To prevent overfitting by randomly dropping units
- C) To enhance positional encoding accuracy
- D) To create more embeddings for training

Answer: B) To prevent overfitting by randomly dropping units

7. What does UMLS BERT specifically aim to enhance?

- A) General language translation tasks
- B) Biomedical natural language processing
- C) Real-time language generation
- D) Image-to-text conversions

Answer: B) Biomedical natural language processing

8. What is the significance of the "scaled dot-product" in transformers?

- A) To reduce computation time
- B) To control the gradient size and stabilize training
- C) To improve output length flexibility

- D) To allow parallel processing in encoders

Answer: B) To control the gradient size and stabilize training

9. Which concept does BioBERT extend from BERT?

- A) Machine translation optimization
- B) Domain-specific training for biomedical texts
- C) Audio processing capabilities
- D) Use of recurrent connections in layers

Answer: B) Domain-specific training for biomedical texts

10. Compared to BERT, which of the following is true about PEGASUS?

- A) PEGASUS uses token-level masking
- B) It is designed primarily for abstractive summarization
- C) PEGASUS applies position embeddings differently
- D) It uses convolutional layers for better text generation

Answer: B) It is designed primarily for abstractive summarization

11. The 'transformer' model primarily replaces which traditional mechanism in NLP?

- A) Fully connected neural networks
- B) Recurrent layers for sequential data processing
- C) Convolutional neural networks
- D) Support vector machines

Answer: B) Recurrent layers for sequential data processing

12. What evaluation framework is associated with BERT for assessing its language understanding?

- A) BLEU score
- B) GLUE benchmark
- C) ROUGE-L
- D) METEOR

Answer: B) GLUE benchmark

13. Which of the following is the primary purpose of segment encoding in transformer models like BERT?

- A) To indicate the position of tokens in a sequence
- B) To differentiate between tokens in different input sentences
- C) To apply dropout to the model for regularization
- D) To scale the dot product in the attention mechanism

Answer: B) To differentiate between tokens in different input sentences

14. If a transformer model overfits when trained on clinical data, which combination of approaches could help mitigate this issue?

- A) Adding more layers
- B) Reducing learning rate
- C) Applying dropout
- D) Using domain-specific pre-training

**Answer: C) Applying dropout
D) Using domain-specific pre-training**

15. Which of the following are benefit(s) of multi-head attention in transformers?

- A) Reduces model training time
- B) Allows the model to focus on different parts of the input simultaneously
- C) Simplifies the transformer architecture
- D) Enhances diverse representation learning

Answer: B) Allows the model to focus on different parts of the input simultaneously

D) Enhances diverse representation learning

16. The use of a _____ in the transformer ensures that each position only attends to earlier positions during training.

Answer: Look ahead mask/ masked attention

17. Which pre-training objectives are used to enhance the capabilities of transformer-based models?

- A) Masked Language Modeling (MLM)
- B) Next Sentence Prediction (NSP)
- C) Gap Sentence Generation (GSG)
- D) Sequence Reordering Objective

Answer: A) Masked Language Modeling (MLM)

B) Next Sentence Prediction (NSP)

C) Gap Sentence Generation (GSG)

18. The _____ mechanism in a transformer model enables it to weigh the importance of different words in an input sequence, helping it focus on relevant parts of the input during processing

Answer: Attention/ self-attention