

## Prova 2 - 2025.2

# Processamento de Linguagem Natural

1. Em um modelo seq2seq recorrente clássico:

- a) A mesma rede recorrente processa a entrada e a saída...
- b) A entrada é analisada em paralelo...
- c) Uma rede processa a sequência de entrada...
- d) A saída é formada diretamente...

2. Modelos seq2seq iniciais, sem atenção, apresentavam o problema de:

- a) Tratar sequências longas...
- b) Condensar toda a informação...
- c) Gerar a sequência de saída...
- d) Manter representações embutidas...

3. O LSTM melhora o RNN simples porque:

- a) Utiliza portas internas...
- b) Introduz uma estrutura de atualização...
- c) Propaga informações apenas...
- d) Distribui o estado recorrente...

4. O papel da atenção em modelos encoder-decoder é:

- a) Permitir que o decoder consulte partes...
- b) Ajustar o estado oculto global...
- c) Propagar informações...
- d) Reorganizar representações...

5. A self-attention viabiliza:

- a) Comparação entre tokens vizinhos...
- b) Imposição de ordem fixa...
- c) Substituição completa...
- d) Interação direta entre todos tokens...

6. Modelos encoder-only são usados em:

- a) Geração longa.
- b) Processamento multimodal.
- c) Classificação.
- d) Tradução direta.

7. Em modelos decoder-only:

- a) Cada token acessa informações globais...
- b) Cada token observa apenas representações anteriores.
- c) Cada token avalia apenas estados iniciais.
- d) Cada token é condicionado exclusivamente...

8. Transformers usam positional encoding porque:

- a) A atenção não contém noção de ordem.
- b) O mecanismo fornece ordem artificial...
- c) Embeddings não podem ser atualizados.
- d) O gradiente depende de posições absolutas.

9. A matriz  $QK^T$  contém:

- a) Similaridades entre tokens.
- b) Apenas valores aleatórios.
- c) Distâncias de posição.
- d) Resultados do feed-forward.

10. No Masked Language Modeling do BERT:

- a) A máscara é prevista com tokens anteriores.
- b) O modelo prevê apenas o próximo token.
- c) A máscara é prevista usando contexto completo.
- d) Treinamento depende de rótulos de sequência.

11. O decoder Transformer contém:

- a) Apenas self-attention...
- b) Self-attention e atenção cruzada...

- c) Uma única projeção linear...
- d) Apenas atenção cruzada.

12. Na etapa de geração seq2seq, usa-se:

- a) Cross-entropy...
- b) Distância euclidiana...
- c) Segmentação hierárquica...
- d) Regressão linear.

13. O vetor de contexto clássico no seq2seq:

- a) É recalculado a cada token.
- b) É único para toda a sequência.
- c) É recalculado apenas no início.
- d) É substituído a cada camada.

14. LSTMs são substituídas em Transformers por:

- a) Atenção.
- b) Pooling.
- c) Softmax.
- d) ReLU.

15. A camada feed-forward no Transformer:

- a) Aplica mesma transformação...
- b) Usa parâmetros diferentes...
- c) Substitui completamente a atenção.
- d) Atualiza apenas embeddings posicionais.

# Gabarito da Prova

## Gabarito

- 1) c
- 2) b
- 3) a
- 4) a
- 5) d
- 6) b
- 7) b
- 8) a
- 9) a
- 10) c
- 11) b
- 12) a
- 13) b
- 14) a
- 15) a