

**QUIZ-II, LLM, B. Tech 7<sup>th</sup> Sem (Sec: 2241016 and Sec: 2241018)**

**Part-1 (Each Question Conveys 1 Mark)**

**1. What is the main function of a tokenizer in Natural Language Processing (NLP) and Deep Learning?**

- A. To generate human-like speech from text
- B. To break down raw text into smaller, manageable units called tokens
- C. To translate text from one language to another
- D. To store large datasets for training models

**2. According to the text, tokenization is described as:**

- A. The last and least important step in NLP
- B. The process of translating human text into images
- C. The first and most crucial step in transforming human language into a numerical format
- D. A method for visualizing language models

**3. In forward propagation, the equation  $z = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$  represents:**

- A. Non-linear transformation
- B. Loss function calculation
- C. Linear transformation in a neuron
- D. Weight update rule

**4. Why is the activation function important in a neural network?**

- A. It performs normalization of input data
- B. It introduces non-linearity to capture complex relationships
- C. It adjusts the learning rate dynamically
- D. It computes the loss value

**5. What is the main innovation introduced by the Transformer model?**

- A. Use of convolutional layers for sequential processing
- B. Replacement of recurrence with self-attention for parallel processing
- C. Use of dropout regularization for translation tasks
- D. Gradient clipping for stable training

**6. What are the two main components of the Transformer architecture?**

- A. Convolution and pooling layers
- B. Input and output gates
- C. Encoder and decoder stacks
- D. Hidden and visible layers

**7. The self-attention mechanism allows each word in a sequence to:**

- A. Attend only to the previous word
- B. Attend to all other words in the sequence to capture dependencies
- C. Ignore distant dependencies in sentences
- D. Be processed strictly in sequential order

**8. Which of the following NLP tasks has been significantly improved using Transformer architectures?**

- A. Sorting algorithms
- B. Question answering and text summarization
- C. Image classification
- D. Signal filtering

**9. Large language models like BERT and GPT are built upon which underlying architecture?**

- A. CNN (Convolutional Neural Network)
- B. RNN (Recurrent Neural Network)
- C. Transformer
- D. LSTM

**10. What happens if the learning rate is too high?**

- A. The model converges very slowly
- B. The model never converges and oscillates around the minimum
- C. The model achieves higher accuracy
- D. The loss decreases smoothly

**11. If the gradient's magnitude is 2.5 and the learning rate is 0.01, what is the parameter change?**

- A. 0.0025
- B. 0.025
- C. 0.25
- D. 0.00025

**12. Batch size refers to:**

- A. The total number of epochs in training
- B. The number of training examples processed before updating weights
- C. The number of neurons in the hidden layer
- D. The size of the output vector

**13. What is an epoch in model training?**

- A. One forward pass through the model
- B. Processing one batch of data
- C. When the model has seen all training examples once
- D. A complete update of all weights and biases

**14. Which of the following correctly defines *Precision*?**

- A. The proportion of actual positives that were correctly identified
- B. The fraction of predicted positives that were actually positive
- C. The harmonic mean of precision and recall
- D. The total number of correct predictions

**15. What does *Recall* measure in a classification model?**

- A. Out of all predicted positives, how many were correct
- B. Out of all actual positives, how many were correctly identified
- C. The ratio between true negatives and false positives
- D. The average of precision and accuracy

**Part-2 (Each Question Conveys 5 Mark) *Answer any Two***

16.

- i. Find the probability of the sentence, “*He reads sea stories daily*” using a bigram language model and the training data (case-insensitive) given below.

Training Sentences:

1. He reads books every evening
2. She reads novels in the morning
3. He sells sea shells by the shore
4. The boy reads a story daily
5. The cat sits by the sea

- ii. What is Token embeddings, and write-down Tokenizer Properties

- iii. How does parallel token processing in Transformer models improve computational efficiency compared to earlier neural network architectures? If possible, explain with figures