

## Introduction to Large Language Models Assignment- 3

Number of questions: 10

Total mark: 10 X 1 = 10

---

### Question 1:

State whether the following statement is True/False.

The Perceptron learning algorithm can solve problems with non-linearly separable data.

- a. True
- b. False

**Correct Answer:** b

**Solution:** The Perceptron algorithm can only handle linearly separable problems.

---

### QUESTION 2:

In backpropagation, which method is used to compute the gradients?

- a. Gradient descent
- b. Chain rule of derivatives
- c. Matrix factorization
- d. Linear regression

**Correct Answer:** b

**Solution:** Backpropagation uses the chain rule of derivatives to calculate the gradients layer by layer.

---

### QUESTION 3:

Which activation function outputs values in the range  $[-1,1]$ ?

- a. ReLU
- b. Tanh
- c. Sigmoid
- d. Linear

**Correct Answer:** b

**Solution:** The tanh function maps input values to the range  $[-1,1]$ .

---

#### QUESTION 4:

What is the primary goal of regularization in machine learning?

- a. To improve the computational efficiency of the model
- b. To reduce overfitting
- c. To increase the number of layers in a network
- d. To minimize the loss function directly

**Correct Answer:** b

**Solution:** As discussed in lecture.

---

#### QUESTION 5:

Which of the following is a regularization technique where we randomly deactivate neurons during training?

- a. Early stopping
- b. L1 regularization
- c. Dropout
- d. Weight decay

**Correct Answer:** c

**Solution:** As discussed in lecture.

---

#### Question 6:

Which activation function has the vanishing gradient problem for large positive or negative inputs?

- a. ReLU
- b. Sigmoid
- c. GELU
- d. Swish

**Correct Answer:** b

**Solution:** The sigmoid function saturates at extreme input values (large positive or negative inputs).

---

#### QUESTION 7:

Which activation function is defined as:  $f(x) = x \cdot \sigma(x)$ , where  $\sigma(x)$  is the sigmoid function?

- a. Swish
- b. ReLU

- c. GELU
- d. SwiGLU

**Correct Answer:** a

**Solution:** As discussed in lecture.

---

#### QUESTION 8:

What does the backpropagation algorithm compute in a neural network?

- a. Loss function value at each epoch
- b. Gradients of the loss function with respect to weights of the network
- c. Activation values of the output layer
- d. Output of each neuron

**Correct Answer:** b

**Solution:** Please refer to the lecture.

---

#### Question 9:

Which type of regularization encourages sparsity in the weights?

- a. L1 regularization
- b. L2 regularization
- c. Dropout
- d. Early stopping

**Correct Answer:** a

**Solution:** L1 regularization encourages sparsity in the weights.

---

#### QUESTION 10:

What is the main purpose of using hidden layers in an MLP?

- a. Helps to the network bigger
- b. Enables us to handle linearly separable data
- c. Learn complex and nonlinear relationships in the data
- d. Minimize the computational complexity

**Correct Answer:** c

**Solution:** Hidden layers enable MLPs to learn complex and nonlinear relationships that a single-layer perceptron cannot model.

---