**Roll No:** | | | | | | | | | | | |

# PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

## Mid Term Examinations - October 2026

**Date:** 07/10/2026 **Time:** 02:00 PM - 03:30 PM

**Course Code:** CAI2502 **Course Name:** Deep Learning

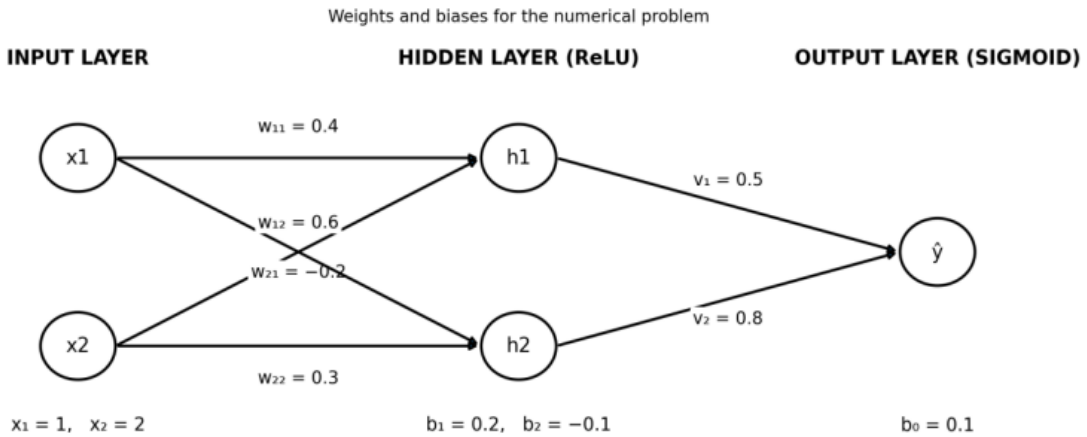
**Semester:** Fifth Semester **Max. Marks:** 50 **Weightage:** 50%

| CO - Levels | CO1 | CO2 |
|---|---|---|
| Marks | 46 | 44 |

### PART A: Answer Following Questions. 10 M

| Q.N. | Questions | M | CO | BT |
|---|---|---|---|---|
| 1 | What is a feedforward neural network? State one application of it. | 2 | CO1 | BT2 |
| 2 | Define underfitting. Mention one method to address it. | 2 | CO1 | BT2 |
| 3 | Differentiate between ReLU and Leaky ReLU activation functions. | 2 | CO1 | BT3 |
| 4 | Why is proper weight initialization important in deep neural networks? | 2 | CO2 | BT2 |
| 5 | What is meant by a learning rate in neural network training? | 2 | CO2 | BT2 |

### PART B : Answer Any 1 Following Questions. 10 M

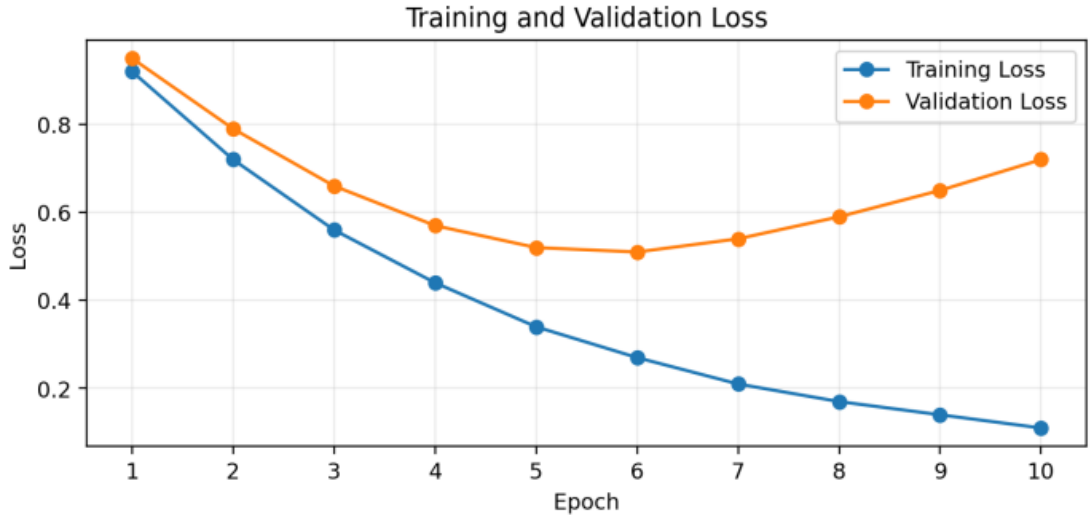
| Q.N. | Questions | M | CO | BT |
|---|---|---|---|---|
| 6 | Explain the working of a Multi-Layer Perceptron (MLP).<br>(a) Explain the function of the input, hidden and output layers. (3)<br>(b) Explain how weights, bias and activation functions determine the output of a neuron. (3)<br>(c) Draw an MLP architecture having 2 input neurons, 3 hidden neurons and 1 output neuron. (4) | 10 | CO1 | BT3 |
| 7 | Consider the neural network shown in Figure 1. The hidden layer uses ReLU activation and the output layer uses sigmoid activation. Given $x_1 = 1$ and $x_2 = 2$, perform the forward pass and determine the predicted output.<br><br>(a) Calculate the net inputs $z_1$ and $z_2$ of the two hidden neurons. (3)<br>(b) Determine $h_1$ and $h_2$ after applying ReLU. (2)<br>(c) Calculate the output-layer net input $z_0$. (2)<br>(d) Calculate $\hat{y}$ using sigmoid, taking sigmoid(1.08) ≈ 0.746. (2)<br>(e) State the predicted class using a threshold of 0.5. (1) | 10 | CO1 | BT3 |

### PART C: Answer Any 1 Following Questions. 10 M

| Q.N. | Questions | M | CO | BT |
|---|---|---|---|---|



PTO (Please Turn Over)

| | | | | |
|---|---|---|---|---|
| 8 | Consider a layer with n-in = 6 and n-out = 3. Answer the following questions on weight initialization.<br>(a) Explain why assigning the same initial value to all weights can prevent effective learning. (2)<br>(b) Calculate the Xavier initialization variance for the given layer. (2)<br>(c) Calculate the He initialization variance for the given layer. (2)<br>(d) Which of the two initializations is more appropriate for a ReLU layer? Give a reason. (2)<br>(e) Explain two possible problems caused by excessively large initial weights. (2) | 10 | CO2 | BT3 |
| 9 | Study Figure 2 and analyze the model's generalization performance.<br><br>(a) Identify the approximate epoch at which overfitting begins. (2)<br>(b) Explain the behavior of training and validation loss after this point. (3)<br>(c) What is meant by the generalization gap? (2)<br>(d) Suggest any two techniques to reduce overfitting. (2)<br>(e) Why is validation loss useful during model training? (1) | 10 | CO2 | BT3 |

| **PART D: Answer Any 1 Following Questions. 10 M** | | | | |
|---|---|---|---|---|
| **Q.N.** | **Questions** | **M** | **CO** | **BT** |
| 10 | A single neuron uses sigmoid activation and Mean Squared Error (MSE) loss. Given x = 3, target y = 0, weight w = 0.4, bias b = −0.2 and learning rate η = 0.1. Use sigmoid(1.0) = 0.731.<br>(a) Calculate the net input and neuron output. (2)<br>(b) Calculate the MSE loss for the given target and predicted output. (2)<br>(c) Using the chain rule, calculate ∂L/∂w. Show all intermediate terms clearly. (4)<br>(d) Update the weight using the calculated gradient and learning rate. (2) | 10 | CO1 | BT3 |
| 11 | A parameter has θ = −1.5, learning rate η = 0.2 and gradient ∂L/∂θ = −0.4. Answer the following.<br>(a) State the Gradient Descent update rule. (2)<br>(b) Calculate the updated value of θ. (2)<br>(c) Compare Batch, Stochastic and Mini-batch Gradient Descent with respect to data used per update and convergence behavior. (4)<br>(d) What may happen if the learning rate is chosen too small? (2) | 10 | CO1 | BT2 |

| **PART E: Answer Any 1 Following Questions. 10 M** | | | | |
|---|---|---|---|---|
| **Q.N.** | **Questions** | **M** | **CO** | **BT** |
| 12 | Explain the following regularization methods and training problems.<br>(a) Explain L1 regularization and its effect on weights. (2)<br>(b) Explain L2 regularization and its effect on weights. (2)<br>(c) Explain Dropout and how it helps reduce overfitting. (2)<br>(d) Explain the purpose of Batch Normalization. (2)<br>(e) What is the vanishing-gradient problem? Mention one suitable remedy. (2) | 10 | CO2 | BT2 |

| | | | | |
|---|---|---|---|---|
| 13 | Compare the following optimization techniques used for training deep neural networks.<br>(a) Explain the role of Momentum in Gradient Descent. (2)<br>(b) Explain Nesterov Accelerated Gradient (NAG). (2)<br>(c) Distinguish between AdaGrad and RMSProp. (2)<br>(d) Explain the main idea of Adam optimization. (2)<br>(e) State the key idea behind AdamW. (2) | 10 | CO2 | BT2 |