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: 06/10/2026 **Time:** 11:45 AM - 01:15 PM

Course Code: CSA2512 **Course Name:** Deep Learning

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

CO - Levels	CO1	CO2	CO3	CO4
Marks	12	32	26	20

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	What is the purpose of an activation function in a neural network?	2	CO1	BT2
2	Given $w=0.8$, learning rate $\eta=0.1$, and gradient $=0.4$, calculate the new weight using one Gradient Descent update.	2	CO2	BT3
3	What is the purpose of a filter/kernel in CNN convolution?	2	CO3	BT2
4	An input image is 8×8 , with a 3×3 filter, stride $=1$, and padding $=0$. Calculate the resulting feature map size.	2	CO3	BT3
5	Perform 2×2 Max Pooling with stride $=2$ on the following feature map: $\begin{bmatrix}2 & 5 & 1 & 3\\4 & 6 & 2 & 1\\7 & 3 & 8 & 4\\2 & 5 & 1 & 9\end{bmatrix}$	2	CO3	BT3

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	Explain Batch Gradient Descent, Mini-Batch Gradient Descent and Stochastic Gradient Descent. Compare them based on the amount of training data used for each update, computational cost and stability of the updates.	10	CO2	BT4
7	Explain the following activation functions with their mathematical expressions and suitable graphs: 1. Sigmoid 2. Tanh 3. ReLU 4. Softmax Compare their characteristics and state their typical use in a neural network.	10	CO1	BT4

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT

PTO (Please Turn Over)

8	Perform convolution on the following *4×4* input matrix using the given *2×2* filter with stride = 1 and padding = 0. Calculate the resulting feature map. Input: $\begin{bmatrix} 1 & 2 & 9 & 4 \\ 5 & 6 & 7 & 10 \\ 11 & 17 & 11 & 32 \\ 13 & 24 & 35 & 26 \end{bmatrix}$ Filter: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	10	CO3	BT3
9	A parameter w is updated using the Gradient Descent rule: $w_{new} = w_{old} - \eta\left(\frac{\partial J}{\partial w}\right)$ The initial value of w = 1.5 and the learning rate η = 0.1. During three successive iterations, the gradients are −0.6, 0.4 and −0.2, respectively. a) Calculate the updated value of w after the first iteration. (3 Marks) b) Calculate the updated value of w after the second iteration. (3 Marks) c) Calculate the updated value of w after the third iteration and explain the effect of the negative gradient on the parameter update. (4 Marks)	10	CO2	BT3

PART D: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
10	Explain the pooling operation used in Convolutional Neural Networks, including Max Pooling and Average Pooling, their effect on feature maps, and the advantages of using pooling layers in CNNs.	10	CO3	BT2
11	Describe the architecture of a feedforward neural network (multilayer perceptron). Explain deep networks, the vanishing/exploding gradient problem, techniques for regularizing a deep network, model exploration and hyperparameter tuning.	10	CO4	BT4

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	Compare and contrast Batch Gradient Descent, Stochastic Gradient Descent (SGD), Momentum, and Adaptive sub-gradient methods such as AdaGrad/RMSProp/Adam. For each method, write the parameter-update equation and discuss advantages and limitations.	10	CO2	BT4
13	Describe the architecture of a feedforward neural network (multilayer perceptron). Explain deep networks, the vanishing/exploding gradient problem, techniques for regularizing a deep network, model exploration and hyperparameter tuning.	10	CO4	BT4