U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

June 2025 Semester End Main Examinations

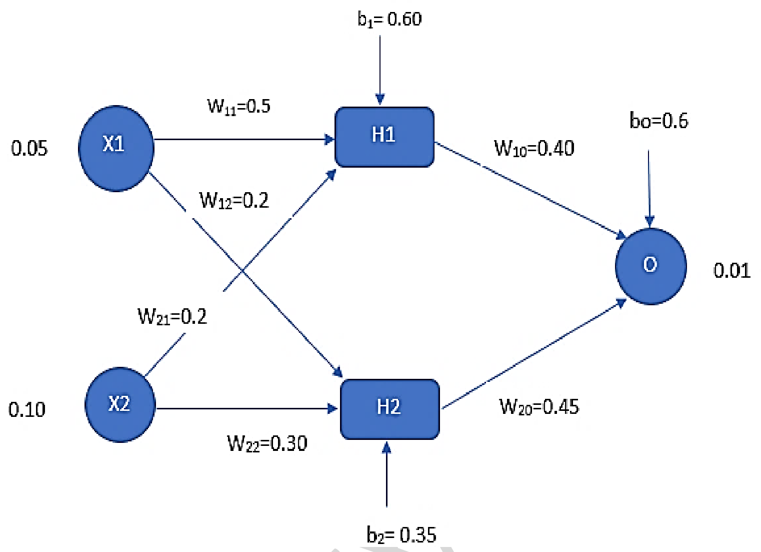
Programme: B.E. **Semester: VII**
Branch: Computer Science and Engineering **Duration: 3 hrs.**
Course Code: 22CS7PENDL **Max Marks: 100**
Course: Neural Network and Deep Learning

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

		UNIT - I	CO	PO	Marks
1	a)	Analyze the behavior of the Sigmoid and Threshold activation functions in the training and performance of the Neural Networks. Justify how does one activation function be preferred over the other.	*CO2*	*PO2*	**5**
	b)	Compare and contrast single-layer feedforward networks, multilayer feedforward networks, and recurrent networks in terms of their structural differences, representational power, and suitability for different types of data and tasks with suitable diagrams.	*CO1*	*PO1*	**7**
	c)	Analyze how the small learning rate parameter affects the bias-variance trade-off in LMS-based learning. How do the natural modes influence the convergence rates of the LMS algorithm?	*CO2*	*PO2*	**8**
		OR			
2	a)	Analyze the functional components of a neuron model, focusing on the synapse, adder, and activation function. Present the nonlinear model of a neuron k in mathematical terms, supplemented by a clear and labeled diagram. Define and explain all the terms involved in the equation.	*CO2*	*PO2*	**5**
	b)	Derive the weight update rule for the perceptron using the perceptron convergence theorem. Develop the error correction learning algorithm for adjusting the weight vector in a basic perceptron model.	*CO2*	*PO2*	**7**
	c)	Examine the Least Mean Square (LMS) algorithm for minimizing the instantaneous cost function. Derive the gradient vector for the steepest descent method and illustrate the process using a signal flow graph that incorporates feedback.	*CO1*	*PO1*	**8**

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

			UNIT - II			
	3	a)	Derive the mathematical formulation of the Universal Approximation Theorem for a nonlinear input-output mapping in neural networks.	***CO2***	***PO2***	**5**
		b)	Consider a multilayer feed-forward neural network given below. Let the learning rate be 0.5. Assume initial values of weights and biases as given in the figure below. Show weight and bias updates by using back-propagation algorithm. Assume that sigmoid activation function is used in the network.	***CO1***	***PO1***	**7**



		c)	Derive the expression for the local gradient $\boldsymbol{\delta_j}(n)$ used to change the synaptic weights in the back propagation algorithm.	***CO2***	***PO2***	**8**
			OR			
	4	a)	Analyse the role of Hessian and its composition of the eigen values, factors effecting the eigen value in the multilayer perceptron trained with the backpropagation algorithm.	***CO2***	***PO2***	**5**
		b)	Show that the supervised learning can be viewed as an optimization problem by summarizing the steps involved in the Nonlinear Conjugate gradient method and Quasi-Newton Method.	***CO2***	***PO2***	**7**
		c)	Derive the expressions for the local gradient of Neuron j when it functions as an output node and when it acts as a hidden node.	***CO2***	***PO2***	**8**
			UNIT - III			
	5	a)	Implement the meta-algorithm using early stopping to determine the best amount of time to train, how long to train and retrain the data, and determine a point at which overfitting begins during training and continue its training until the value is reached.	***CO3***	***PO3***	**10**

		b)	Analyze the role of dropout in neural networks in mitigate overfitting. What are the computational trade-offs and the overall impact of dropout on model generalization.	***CO2***	***PO2***	**5**
		c)	Analyze the impact of applying aggressive data augmentation on the quality of the learned features.	***CO2***	***PO2***	**5**
			OR			
	6	a)	Explain how parameter norm penalties act as a regularization technique. Derive the mathematical expression for L2 parameter regularization and its incorporation into the loss function.	***CO1***	***PO1***	**10**
		b)	Apply parameter tying and sharing on the optimization landscape and model convergence.	***CO1***	***PO1***	**5**
		c)	Analyze how multitask learning improves the generalization ability of models compared to single-task learning.	***CO2***	***PO2***	**5**
			UNIT - IV			
	7	a)	Show the different variants of the convolution in terms of convolution with a stride and the effect of zero padding on network size with appropriate diagrams.	***CO1***	***PO1***	**10**
		b)	i) For the input matrix: $$\begin{bmatrix} 10 & 20 & 30 & 40 \\ 15 & 25 & 35 & 45 \\ 20 & 30 & 40 & 50 \\ 25 & 35 & 45 & 55 \end{bmatrix}$$ A. Perform **max pooling** with a 2×2 kernel and stride 2. B. Perform **average pooling** with a 2×2 kernel and stride. Compare the two resulting matrices. ii) Discuss the three stages involved in the convolution network, how does the third stage of applying the pooling function modify the output layer.	***CO1***	***PO1***	**10**
			OR			
	8	a)	i) With a neat diagram, demonstrate the convolution variations for the locally connected layers, tiled convolution and standard convolution. ii) Explain the three special cases of zero padding settings in terms of width of the image and the width of the kernel.	***CO1***	***PO1***	**10**
		b)	i) Illustrate the different formats of data that can be used with convolutional networks. ii) Justify that the kernel that is separable is inefficient.	***CO1***	***PO1***	**10**

			UNIT - V			
	9	a)	Design the recurrent neural networks with different patterns that map an input sequence of x values to a corresponding sequence of output o values and the measure of loss L to calculate the difference of o from training target y and explain the same.	***CO3***	***PO3***	**10**
		b)	Any model representing a variable P (y; θ) can be reinterpreted as a model representing a conditional distribution P (y|ω) with ω = θ. We can extend such a model to represent a distribution P(y | x) by using the same P(y | ω) as before, Design the different types of RNNs that generates the y sequence by considering extra input.	***CO3***	***PO3***	**10**
			OR			
	10	a)	Design the gradient term for the output o(t) and hidden layer h(t) using the Back Propagation through time (BPTT) algorithm in the Recurrent Neural Network.	***CO3***	***PO3***	**10**
		b)	i) Obtain the gradients on the parameters bias vectors b and c, weight matrix U (input -to -hidden),V (hidden- to- output), W (hidden-to-hidden) of the computational graph. ii) Design an encoder- decoder sequence architecture that can map the input sequence to an output sequence that are of different lengths. Also describe the process of the mapping using the architecture.	***CO2,3***	***PO2,3***	**10**

******