U.S.N.										

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

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: VI**

Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**

Course Code: 24AM6PCDEL **Max Marks: 100**

Course: 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.

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 - I	CO	PO	Marks
1	a)	Illustrate how ReLU and tanh activation functions exhibit saturation.	CO2	PO2	**06**
	b)	Consider a Deep Neural Network (DNN) with the following structure: • Inputs to the hidden layer: z = [0.8, 0.6, 0.7, 0.9]. • Weight matrix connecting the hidden layer to the output layer: 0.2 0.5 0.3 0.8 0.7 0.1 0.6 0.4 0.9 0.2 0.5 0.3 • Bias vector for the output layer: [0.1, 0.3, 0.5] • Activation function at the hidden layer: ReLU • Activation function at the output layer: Softmax Compute the output probabilities for each class.	CO3	PO3	**10**
	c)	Describe any two heuristics that can be used to accelerate the training process of a DNN.	CO1	PO1	**04**
		OR			
2	a)	Describe the significance of injecting noise at the output targets	CO2	PO2	**06**
	b)	Consider a binary classification problem where a neural network predicts whether an image contains a cat (class 1) or not (class 0). True labels are y = [1, 0, 1, 1, 0] and the predicted probabilities for class 1 are p = [0.9, 0.3, 0.8, 0.88, 0.1]. Apply cross-entropy loss function and compute the error between the predicted probabilities and the true labels.	CO3	PO3	**10**
	c)	Explain the need of data augmentation.	CO1	PO1	**04**
		UNIT - II			
3	a)	Write an early stopping meta-algorithm to determine the best amount of training time.	CO1	PO1	**06**

	b)	How does gradient clipping help in addressing the exploding gradient problem?	*CO1*	*PO1*	**07**
	c)	Consider a densely connected base network featuring two visible units, one hidden layer containing two hidden units, and a single output unit. Illustrate the base network and depict all possible sub-networks that can be generated by dropping out various subsets of units from the base network.	*CO2*	*PO2*	**07**
		OR			
4	a)	Explain parameter tying and parameter sharing using a suitable application.	CO1	PO1	**06**
	b)	Differentiate RMSProp and RMSProp with Nesterov momentum.	CO1	PO1	**06**
	c)	Explain the use of early stopping as a meta-algorithm to i. Identify the onset of overfitting during training. ii. Determine the optimal number of training steps for retraining on the full dataset.	CO1	PO1	**08**
		UNIT - III			
5	a)	Using suitable mathematical equations differentiate between Long Short Term Memory (LSTM) and Gated Recurrent Units (GRUs).	*CO1*	*PO1*	**06**
	b)	Compute the values of input gate and output gate of a LSTM cell by considering the given data Weights for the input gate: $W_i = \begin{bmatrix} 0.5 & 0.4 \\ 0.3 & 0.2 \end{bmatrix}$ Hidden state: $h_{t-1} = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$ Recurrent weights for the input gate: $U_i = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix}$ Biases for the input gate: $b_i = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$ Input vector: $x_t = \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix}$ Weights for the output gate: $W_o = \begin{bmatrix} 0.7 & 0.6 \\ 0.5 & 0.4 \end{bmatrix}$ Recurrent weights for the output gate: $U_o = \begin{bmatrix} 0.3 & 0.4 \\ 0.5 & 0.6 \end{bmatrix}$ Biases for the output gate: $b_o = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$	*CO3*	*PO3*	**10**
	c)	Explain the vanishing gradient problem in Recurrent Neural Network (RNN).	*CO1*	*PO1*	**04**
		OR			
6	a)	How does Vanilla RNN differ from Bi-directional RNN?	*CO2*	*PO2*	**06**
	b)	Compute the values of update gate and the Candidate Hidden State $\tilde{h}_t$ of a GRU cell by considering the given data Weights for the update gate: $W_z = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.5 & 0.6 & 0.7 & 0.8 \end{bmatrix}$ Biases for the update gate: $b_z = [0.1, 0.2]$ Input vector	*CO3*	*PO3*	**10**

Previous hidden state $h_{t-1} = [0.1, 0.4]$	$x_t = [0.5, -0.2]$
Candidate Activation $W_h = \begin{bmatrix} 0.1 & 0.2 & 0.3 & 0.4 \\ 0.5 & 0.6 & 0.7 & 0.8 \end{bmatrix}$ $b_h = [0.1, 0.2]$	Reset Gate $r_t \approx \begin{bmatrix} 0.5646 \\ 0.6630 \end{bmatrix}$

c) Write the Back Propagation Through Time (BPTT) algorithm. — *CO1* | *PO1* | **04**

UNIT - IV

7 a) Consider a Convolutional Neural Network (CNN) with the following configuration: — *CO3* | *PO3* | **10**

An input image of size 4x4:

1	2	0	3
4	6	2	1
1	2	5	3
0	1	2	4

A single 2x2 filter (kernel) with weights:

0	1
-1	1

i. Perform convolution operation on the given input image using the specified filter and stride = 1. Write the resulting feature map.

ii. Apply ReLU activation function to the result obtained in (i). Represent the updated feature map in the matrix form.

iii. Perform max pooling on the feature map obtained from (ii) using 2*2 pooling window and stride = 1. Determine the resulting feature map.

b) Is it possible to retain the original size of the image after the process of convolution? Justify with an example. — *CO2* | *PO2* | **06**

c) Elaborate on key application areas of adversarial training. — *CO1* | *PO1* | **04**

OR

8 a) Design a CNN with the following specification to classify images of handwritten digits between 0 and 9. Represent the size of the intermediate feature maps precisely on the architectural design. — *CO3* | *PO3* | **10**

- **Input**: (28, 28, 1)
- **Conv1**: 32 filters, kernel: 3x3, activation function: ReLU
- **Pool1**: 2x2 Max Pooling
- **Conv2**: 64 filters, kernel: 3x3, activation function: ReLU
- **Pool2**: 2x2 Max Pooling
- **Flatten**
- **Dense**: 128 neurons, activation function: ReLU
- **Output**: 10 neurons, activation function: Softmax

b) Given an image of size 30*30 with 3*3 kernel, horizontal stride = 2 and vertical stride = 2. Calculate the size of the resultant feature map. — *CO3* | *PO3* | **04**

		c)	Can Manifold Tangent Classifier be used to improve the accuracy of classification? Justify.	*CO2*	*PO2*	**06**
			UNIT - V			
	9	a)	Explain process of training a Hopfield network along with relevant mathematical representations.	*CO1*	*PO1*	**07**
		b)	Illustrate the steps of Contrastive divergence algorithm, for training Restricted Boltzmann Machine (RBM).	*CO2*	*PO2*	**05**
		c)	Consider an RBM with 2 visible units and 2 hidden units. Given the weights and biases: Weight matrix $W = \begin{bmatrix} 0.5 & -0.7 \\ 0.2 & 0.3 \end{bmatrix}$ Visible unit biases b_v: [0.1, -0.2] Hidden unit biases b_h: [0.3, 0.4] Calculate the probabilities of the hidden units being activated given a specific visible vector v=[1,0].	*CO3*	*PO3*	**08**
			OR			
	10	a)	Illustrate the training process of multilayer Restricted Boltzmann Machines.	*CO2*	*PO1*	**07**
		b)	Differentiate between Boltzman Machine and Restricted Boltzman Machine.	*CO2*	*PO2*	**05**
		c)	A Restricted Boltzmann Machine (RBM) with 3 visible units and 2 hidden units has the following initial weights W between the visible and hidden units: [0.1, -0.2; 0.4, 0.5; -0.3, 0.2] The biases for the visible units and the hidden units are: [0.1; 0.2; -0.1] [0.1; -0.1] Given an input vector $v = (1\ 0\ 1)$ and learning rate = 0.1 use sigmoid function and illustrate the process of weight updates using Contrastive Divergence algorithm for 1 epoch.	*CO3*	*PO3*	**08**

******