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: VI**
Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**
Course Code: 24AM6PCDEL / 22AM6PCDEL **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)	Write the Stochastic Gradient Descent (SGD) with Nesterov momentum algorithm.	*CO1*	*PO1*	**06**
	b)	Using mathematical equations, demonstrate how both linear and non-linear variants of the ReLU activation function will address the "dying ReLU" problem.	*CO2*	*PO2*	**10**
	c)	Illustrate the significance of data augmentation.	*CO1*	*PO1*	**04**
		OR			
2	a)	Explain the role of ReLU properties in preventing local minima in deep neural networks.	*CO1*	*PO1*	**06**
	b)	i. Describe the challenges posed by the vanishing gradient problem in deep learning models. ii. Discuss strategies to address this issue.	*CO1*	*PO1*	**06**
	c)	For a binary classification problem where a neural network predicts whether an image contains a cat (class 1) or not (class 0), with true labels y = [0,0,1,1,1] and predicted probabilities p=[0.8,0.4,0.9,0.77,0.2], calculate the cross-entropy loss to determine the error between the predictions and the true labels.	*CO2*	*PO1*	**08**
		UNIT - II			
3	a)	Explain the impact of early stopping and detail on its role as a regularization technique.	*CO1*	*PO1*	**08**
	b)	Elaborate on the limitations of using gradient clipping.	*CO2*	*PO2*	**06**
	c)	Describe the concept of parameter tying and parameter sharing using a suitable application.	*CO1*	*PO1*	**06**
		OR			
4	a)	Compare RMSProp and RMSProp with Nesterov momentum.	*CO1*	*PO1*	**06**
	b)	Consider a fully connected base network with two input units, one hidden layer containing two hidden units, and one output unit.	*CO2*	*PO1*	**08**

		Illustrate the base network and show all possible sub-networks that can be created by applying dropout to various subsets of the units.			
	c)	Write an early stopping meta-algorithm to determine the best amount of training time.	*CO2*	*PO1*	**06**
		UNIT - III			
5	a)	Compute the values of forget gate and new cell state $\tilde{c}_t$ of a Long Short Term Memory (LSTM) cell by considering the given data Weights for the forget gate $W_f = \begin{bmatrix}0.6 & 0.5\\0.4 & 0.3\end{bmatrix}$ Hidden state $h_{t-1} = \begin{bmatrix}0.1\\0.2\end{bmatrix}$ Recurrent weights for the forget gate $U_f = \begin{bmatrix}0.2 & 0.3\\0.4 & 0.5\end{bmatrix}$ Biases for the forget gate $b_f = \begin{bmatrix}0.1\\0.2\end{bmatrix}$ Input vector $x_t = \begin{bmatrix}0.5\\0.6\end{bmatrix}$ Weights for the cell state $W_c = \begin{bmatrix}0.2 & 0.1\\0.1 & 0.2\end{bmatrix}$ Recurrent weights for the cell state $U_c = \begin{bmatrix}0.4 & 0.5\\0.6 & 0.7\end{bmatrix}$ Biases for the cell state $b_c = \begin{bmatrix}0.1\\0.2\end{bmatrix}$	*CO3*	*PO3*	**10**
	b)	Differentiate between Bi-directional Recurrent Neural Networks (RNN) and Multi-layer Recurrent Neural Networks.	*CO2*	*PO2*	**06**
	c)	Elaborate on the exploding gradient problem in Recurrent Neural Network.	*CO1*	*PO1*	**04**
		OR			
6	a)	Compute the values of reset gate and the new hidden State h_t of a Gated Recurrent Unit (GRU) cell by considering the given data Weights for the reset gate $W_r = \begin{bmatrix}0.1 & 0.2 & 0.3 & 0.4\\0.5 & 0.6 & 0.7 & 0.8\end{bmatrix}$ Previous hidden state $h_{t-1} = [0.1, 0.4]$ Biases for the reset gate $b_r = [0.1, 0.2]$ Input vector $x_t = [0.5, -0.2]$ Candidate hidden state $\tilde{h}_t \approx \begin{bmatrix}0.2246\\0.5208\end{bmatrix}$ Update Gate $z_t \approx \begin{bmatrix}0.5646\\0.6630\end{bmatrix}$	*CO3*	*PO3*	**10**
	b)	Illustrate the time layered representation of RNN.	*CO2*	*PO2*	**06**
	c)	Describe the modified Back Propagation Through Time (BPTT) algorithm used in bidirectional RNN.	*CO1*	*PO1*	**04**
		UNIT - IV			
7	a)	Design a Convolutional Neural Network (CNN) with the following specifications to classify images of cats and dogs.	*CO3*	*PO3*	**10**

<table>
<tr><td></td><td></td><td>Represent the size of the intermediate feature maps precisely on the architectural design.

- **Input**: (32, 32, 3 channels)
- **Conv1**: 32 filters, kernel: 3x3, padding = 1, activation function: ReLU
- **Pool1**: 2x2 Max Pooling, Stride = 2
- **Conv2**: 64 filters, kernel: 3x3, padding = 1, activation function: ReLU
- **Pool2**: 2x2 Max Pooling , Stride = 2
- **Conv3**: 128 filters, kernel: 3x3, padding = 1, activation function: ReLU
- **Pool3**: 2x2 Max Pooling , Stride = 2
- **Flatten**
- **Dense**: 256 neurons, activation function: ReLU
- **Output**: 2 neurons, activation function: Softmax
</td><td></td><td></td><td></td></tr>
<tr><td></td><td>b)</td><td>Explain the significance of pooling in CNN.</td><td>*CO1*</td><td>*PO1*</td><td>**04**</td></tr>
<tr><td></td><td>c)</td><td>How does Tangent Propagation help in reducing over fitting of a model?</td><td>*CO2*</td><td>*PO2*</td><td>**06**</td></tr>
<tr><td></td><td></td><td>**OR**</td><td></td><td></td><td></td></tr>
<tr><td>8</td><td>a)</td><td>Consider a Convolutional Neural Network (CNN) with the following configuration:
An input image of size 6x6:
<table>
<tr><td>1</td><td>2</td><td>3</td><td>0</td><td>1</td><td>2</td></tr>
<tr><td>4</td><td>5</td><td>6</td><td>1</td><td>0</td><td>3</td></tr>
<tr><td>7</td><td>8</td><td>9</td><td>2</td><td>1</td><td>0</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>0</td><td>1</td><td>2</td></tr>
<tr><td>4</td><td>5</td><td>6</td><td>1</td><td>0</td><td>3</td></tr>
<tr><td>7</td><td>8</td><td>9</td><td>2</td><td>1</td><td>0</td></tr>
</table>
Two 3x3 filter (kernel) with weights:
<table>
<tr><td>1</td><td>0</td><td>-1</td></tr>
<tr><td>1</td><td>0</td><td>-1</td></tr>
<tr><td>1</td><td>0</td><td>-1</td></tr>
</table>
<table>
<tr><td>1</td><td>1</td><td>1</td></tr>
<tr><td>0</td><td>0</td><td>0</td></tr>
<tr><td>-1</td><td>-1</td><td>-1</td></tr>
</table>

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 = 2. Determine the resulting feature map.
</td><td>*CO3*</td><td>*PO3*</td><td>**10**</td></tr>
<tr><td></td><td>b)</td><td>Illustrate the effect of increasing the stride on the spatial dimensions of the output feature map in a convolutional layer.</td><td>*CO2*</td><td>*PO2*</td><td>**04**</td></tr>
<tr><td></td><td>c)</td><td>Differentiate between Tangent distance, Euclidian distance and Manifold distance.</td><td>*CO1*</td><td>*PO1*</td><td>**06**</td></tr>
</table>

			UNIT - V			
	9	a)	Elaborate on the process of learning the weights of a Boltzmann Machine using appropriate mathematical representations.	*CO2*	*PO2*	**07**
		b)	A Boltzmann Machine consists of two visible units v_1, v_2 and a hidden unit h. The weights between these units are as follows: Weight between v_1 and h = 1, Weight between v_2 and h = -1. There are no direct connections between v_1 and v_2. The states of the units can be either 0 or 1. Calculate the energy E(v_1, v_2, h) of the system for the state (v_1, v_2, h) = (1, 0, 1)	*CO3*	*PO3*	**05**
		c)	Describe the procedure involved in training a multilayer Restricted Boltzmann Machine (RBM).	*CO2*	*PO2*	**08**
			OR			
	10	a)	Describe the process of training a Hopfield network along with relevant mathematical representations.	*CO1*	*PO1*	**08**
		b)	Explain the challenges in applying RBM's to non-binary data compared to binary data?	*CO1*	*PO1*	**06**
		c)	Differentiate Restricted Boltzman Machine and Boltzman Machine.	*CO1*	*PO1*	**06**

******