1511E1274

Candidate's Seat No :_____

Int. M.Sc. (DS) Sem.-7 Examination

CC-403

Deep Learning

November-2025

Time : 2-30 Hours] [Max. Marks : 70

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

Q.1 **(a)** Define deep learning and explain how it differs from traditional machine learning approaches. **(07)**

(b) What are the different types of activation functions popularly used? Explain each of them. **(07)**

OR

(a) Explain in detail the basic structure of Perceptron. **(07)**

(b) Describe the backpropagation algorithm and its significance in training ANNs. **(07)**

Q.2 **(a)** Explain, in detail, the basic structure of Multi-Layer Perceptron. **(07)**

(b) A neuron has two binary inputs and uses a threshold activation function. Its weights are $w_1 = 2$ and $w_2 = 1$, and the bias is $b = -2.5$. Determine the output of the neuron for each of the following inputs: (0, 0), (1, 0), (0, 1), and (1, 1). Also, identify the logical function (if any) this neuron is computing. **(07)**

OR

(a) Explain the working of dropout, including their mathematical operations and regularizing effects. **(07)**

(b) What is the vanishing and exploding gradient problem? Provide an example where the gradients vanish for the sigmoid activation function. **(07)**

Q.3 **(a)** Describe the theoretical and practical impact of batch normalization, including their mathematical formulations and roles in model training. **(07)**

(b) A Convolutional Neural Network (CNN) receives an input image of size 28×28×3. The network is composed of the following layers: **(07)**

a. Convolutional Layer 1: 16 filters of size 5×5, stride 1, padding 2
b. Max Pooling Layer 1: 2×2 window, stride 2
c. Convolutional Layer 2: 32 filters of size 3×3, stride 1, padding 1
d. Max Pooling Layer 2: 2×2 window, stride 2
e. Fully Connected Layer: 256 neurons
f. Output Layer: 10 neurons for classification

Answer the following:

i. Calculate the output dimensions (height × width × depth) after each layer.
ii. Calculate the total number of trainable parameters (weights + biases) in each convolutional and fully connected layer.
iii. What is the total number of trainable parameters in the entire network?

OR

(a) Compare and contrast Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) in terms of their architectural characteristics, methods of processing data, and common application areas. **(07)**

(b) Discuss the concept of receptive fields in CNNs. How does it relate to feature extraction and spatial hierarchy in image data? **(07)**

P.T.O

E1274-2

Q.4 (a) Explain the various types of sequence modeling based on input and output sequence structures. Include relevant examples for each category. **(07)**

(b) Explain the working of a vanilla Recurrent Neural Network (RNN) and describe its layer structure. Use a neat and labeled diagram to illustrate how it maintains memory of previous inputs over time **(07)**

OR

(a) Discuss the strengths and limitations of Recurrent Neural Networks (RNNs) in handling sequential data. **(07)**

(b) What are Gated Recurrent Units (GRUs)? Explain the architecture of a GRU cell with the help of a labeled diagram. **(07)**

Q.5 Attempt any **SEVEN** out of **TWELVE:** **(14)**

(1) What is the purpose of the bias input in a Perceptron, and why is it necessary?

(2) Compute the total number parameters (Weights and bias) when Layer sizes: 12, 14, 14, 14, 2.

(3) What is the key idea behind Momentum optimization?

(4) How does AdaGrad differ from standard gradient descent?

(5) What makes Adam the most popular optimizer in deep learning?

(6) Why is padding commonly used in convolutional layers, and what is its general effect on the output dimensions?

(7) Describe the concept of "weight sharing" in convolutional layers and explain its main benefits.

(8) What problem did ResNet primarily address, and what architectural innovation did it introduce to solve it?

(9) Distinguish between Max Pooling and Average Pooling in terms of their operation and typical use cases.

(10) Distinguish between sequence data and time series data.

(11) What distinguishes a Gated Recurrent Unit (GRU) from a Long Short-Term Memory (LSTM) network?

(12) Illustrate vanishing gradient problem in RNN.

**** ****