[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 7966 **L**

Unique Paper Code : 2343010017

Name of the Paper : Deep Learning

Name of the Course : **DSE**

Semester : IV/VI/VIII

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Question No. **1 (Section A)** is compulsory.
3. Attempt any **4 (four)** questions from **Section B**.
4. Parts of a question should be attemptedtogether.
5. Use of scientific calculator is allowed.

Section A

1. (a) Explain RMSProp and ADAM optimizer. Also give two advantages and two disadvantages of each. (6)

   (b) How does the design of an LSTM cell address the vanishing gradient problem found in traditional RNNs, and what specific roles do the forget, input, and output gates play in this process? (6)

P.T.O.

(c) Explain the role of activation functions in neural networks. Write the mathematical expressions for Sigmoid, ReLU, and Linear activation functions. Also compute the output of each function for the input x=2.

(d) Consider a convolutional neural network (CNN) with the following specifications: (5)

Input size: 28×28×3(RGB image)

Number of filters: 6, Filter size: 5×5, Stride: 1, Padding: 0

(i) Calculate the total number of trainable parameters in this convolutional layer.

(ii) Determine the dimensions of the output feature map produced by this layer.

(e) What is data augmentation in deep learning? A dataset contains 800 images. Using data augmentation, each original image is rotated, flipped andcropped producing 3 additional images. Find the total number of images after augmentation. (3)

(f) Given: (4)

$y = 0.8, t = 1$

Input x= 2, learning rate η=0.1

Initial weight w=0.

Find the gradient $\frac{\partial E}{\partial w}$ using:

$$\frac{\partial E}{\partial w} = (y - t).x$$

Find the updated weight after one step.

Section B

2. (a) Explain the architecture and working of a Generative Adversarial Network (GAN). Describe the roles of the generator and discriminator and explain how the network is trained through the adversarial learning process. Also mention one practical application of GANS. (7)

(b) Explain the concept of regularization in neural networks. Then, compare the L^1 and L^2 regularization techniques, highlighting their differences and effects on model training. (5)

(c) Explain the intuition behind using momentum in gradient descent. (3)

3. (a) Describe the working of batch normalization in deep learning. List advantages of batch normalization. (6)

(b) A neural network layer has 10 neurons. During training, a dropout rate of (3) 0.3 is applied.

Find:

(i) Number of neurons dropped

(ii) Number of neurons remaining

(iii) 1f each active neuron has output value 2, find the total output after dropout (3)

(c) Writeany three challenges in neural network optimization? (3)

(d) Given the following sentence: (3)

"This is deep learning course paper."

Construct the unigrams, bigrams, and trigrams(3-grams) for the sentence.

4. (a) Design a single perceptron to implement the AND logic gate. Draw the diagram, specify the weights and bias, and demonstrate its operation by showing the outputs for all binary input combinations (00, 01, 10, 11). (6)

(b) Draw the unfolded representation of a simple RNN for three time steps (t = 3). Also, write the equation that defines the current hidden state of the RNN. (5)

(c) Distinguish between sparse autoencoders and denoising autoencoders. (4)

5. (a) Perform a 2D convolution on the following input matrix using the given kernel, with stride = 1 and no padding. (6)

Input Matrix (6 × 6):

1	2	0	1	2	1
0	1	3	2	1	0
2	1	0	1	2	1
1	2	1	3	0	2
0	1	2	1	1	0
1	0	1	2	0	1

Kernel (3×3)

0	1	0
1	1	1
0	1	0

(i) Compute the feature map after convolution

(ii) Apply the ReLU activation function

(iii) Perform max pooling using a (2 x 2) window with stride = 1

(b) Define epoch, batch size, and iteration. For a dataset with number of samples = 500, batch size = 50 andepochs = 4, compute the number ofiterations per epoch and the total number of iterations. (5)

(c) Explain the "Bias-Variance Trade-off" and discuss how model complexity influences both. Support your answer with a description of the relationship between model complexity and total error. (4)

6. For a simple autoencoder given below:

Input Vector:x= [2,1,4]

Encoder Weights:w= [0.3,0.5,0.2], bias b = 0.4

Decoder Weights:w= [0.6,0.2,0.7], bias b = 0.2

(i) Calculate the encoded representation of the input.

(ii) Calculate the Mean Squared Error (MSE) between the original and reconstructed input.

(iii) Draw a diagram of the autoencoder.

(b) Explain the difference between Global Contrast Normalization and Local Contrast Normalization. (5)

(c) Given an LSTM layer with an input dimension of 5 and 3 hidden units, determine the total number of trainable parameters in the layer. (3)

7. (a) Consider a neural network with:

- Input layer: 2 inputs $x_1 = 1$; $x_2 = 2$
- Hidden layer: 2 neurons (h_1, h_2), both using ReLU activation
- Output layer: 1 neuron with linear activation

The weights and biases are as follows: -

Input to Hidden Layer:

Neuron h_1:$W_{11} = 0.3$;W_{21}=0.2; $b_1 = 0.1$

Neuron h_2:W_{12}=-0.2; $w_{22} = 0.5$; $b_2 = 0.3$

Hidden to Output Layer:$w_{10} = 0.4$; $W_{20} = 0.6$; b_0 0.2

Assume the input is: $x_1 = 1$; $x_2 = 2$;

(i) Draw the neural network structure for above and perform a forward pass:

a. Compute h_1,$h_2$using ReLU

b. Compute the final output (y)

(ii) Given target value: Y_{target}=2, compute the Mean Squared Error (MSE):

(b) Describe the mathematical process of a 2D Convolution operation. How do 'stride' and 'padding' affect the output? (5)

(500)