[This question paper contains 4 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 272 **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 attempted together.
5. Use of scientific calculator is allowed.

Section A

Q1 (a) Identify and explain the optimization algorithm that integrates the concepts of momentum and adaptive learning rates. Provide two distinct advantages and two disadvantages of utilizing this algorithm in deep learning models. (6)

(b) What is Batch Normalization? Explain its purpose and describe the steps involved in performing Batch Normalization during training. (5)

(c) Consider a neural network with: (4)

One hidden layer containing two neurons (h_1 and h_2), both using ReLU activation andone output layer with one neuron using linear activation.

The weights and biases are given 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) Perform a forward pass and compute the output of the network

ii) If the target output is 2, calculate the Mean Squared Error (MSE)

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

(e) For a CNN layer with input size 28×28×1 (Grayscale image), number of filters = 16, filter size = 5×5, stride = 1, and no padding, determine: (4)

i) The number of trainable parameters in the layer

ii) The size of the resulting feature map

(f) What is data augmentation? List any two areas where it is applied. (4)

(g) Define weight sparsity and name the regularization method used to achieve it. Why is this method particularly useful when dealing with high-dimensional datasets containing redundant features? (3)

Section B

Q2 (a) Explain the working of the backpropagation algorithm. Describe the chain rule of calculus and its role in backpropagation. (8)

(b) What is dropout? Explain how dropout improves the generalization ability of deep learning models. Discuss the advantages and limitations of dropout in deep neural networks. (4)

(c) Identify and explain two key limitations of n-gram models that are effectively mitigated using word embeddings in neural network architectures. (3)

Q3 (a) In the context of CNN, an input grayscale image of dimension (5×5) is given as: (9)

2	-1	0	2	1
1	3	2	0	1
0	1	4	1	2
2	0	1	3	1
1	2	0	1	2

Perform the convolution operation with the following kernel of dimension (3×3),

0	1	0
1	1	1
0	1	0

With stride = 1, and no padding:

(b) On the feature map obtained in Q3 a) above, apply the ReLU activation function and then perform max pooling with a 2×2window and stride = 1. (3)

(c) A weight has an initial value of 0.8, and the gradient of the loss with respect to this weight is 0.5. Given a learning rate of 0.05, find the new value of the weight after one step of gradient descent. (3)

Q4 (a) Discuss the paradigm shift in speech recognition that occurred when Deep Neural Networks (DNNs) replaced Gaussian Mixture Models (GMMs). What specific limitations of GMM-HMM systems did DNNs successfully address? (6)

(b) Describe the importance of activation functions in neural networks. (5)

Write the mathematical expressions for the following activation functions:

(i) Sigmoid

(ii) ReLU

(iii) Linear

Also, calculate the output of each activation function when the input value is x is4.

(c) An autoencoder is designed for image compression with the following architecture: (4)

Input Layer: 1,024 units (representing a 32×32pixel image)

Bottleneck Layer: 128 units

Output Layer: 1,024 units

Assuming the network is fully connected and there are no bias terms, calculate the total number of trainable parameters.

Q5 (a) Describe the architecture of an LSTM cell with the help of a clear diagram. (6)

Explain the roles of the input gate, forget gate, output gate, and the cell state.

(b) Explain why saddle points and flat regions (plateaus) slow down the optimization process during the training of neural networks. (5)

(c) Consider a simple recurrent neural network (RNN) where the hidden state is updated using the rule:$h_t = \tanh(0.5\, x_t + 0.4\, h_{t-1})$ (4)

Here, x_t represents the input at time step t,

and h_t is the hidden state at time t.

Given the input sequence x=[1,2] and the initial hidden state h_o=0 compute the hidden states , h_1 and h_2.

Q6 (a) Show that a single-layer perceptron cannot realize the XNOR function. (6)
Construct a two-layer neural network with step activation that implements XNOR, specifying the weights and biases for each hidden neuron and for the output neuron, along with the activation functions used.

(b) In a GAN, the discriminator outputs the probability that an input is real. For a batch of samples, suppose: (5)

- The discriminator outputs D(x) = 0.9 for a real sample
- The discriminator outputs D(G(z)) = 0.3 for a generated (fake) sample

Using the GAN value function,

$$L_C = \log D(x) + \log(1 - D(G(z)))$$

calculate the total loss for the discriminator.

(c) Discuss one application each of a deep neural network with respect to (4)

- Computer vision
- Speech Recognition

Q7 Write short notes on any five of following: (15)

(a) Bidirectional RNN model
(b) Global Contrast Normalization
(c) Graphics Processing Unit
(d) Hierarchical Softmax
(e) Ill-Conditioned Hessian Matrix
(f) Vanishing/Exploding gradient problem