Printed Pages– 7 Roll No. ....

D127871(022)

B. Tech. (Hon's) (Eighth Semester)
Examination, April-May 2025

(Computer Science and Engineering - Artificial Intelligence/Data Science Branch)

DEEP LEARNING

Time Allowed : Three hours

Maximum Marks : 100

Minimum Pass Marks : 35

Note *: Each question four parts. Part (a) of each question in compulsory. Attempt any* ***two*** *parts from (b), (c) and (d) of each question. Part (a) is of 4 marks and part (b), (c) and (d) has 8 marks. The figure in the right-hand margin indicates marks.*

Unit-I

1. (a) Exlain the Multivariate Chain Rule for computing gradients in a neural network. Illustrate with a

concrete numerical example involving a small feed-forward network (one hidden neuron) where

$$y = f(z),\ z = w_1x_1 + w_2x_2,$$

$$f(z) = \sigma(z) = \frac{1}{1+e^{-z}}$$

compute $\frac{\partial L}{\partial w_1}$ if $L = \left(y_{pred} - y_{true}\right)^2$, given

$x_1 = 1, x_2 = 2, w_1 = 0.2, w_2 = -0.1, y_{true} = 1$ 4

(b) Give an overview of the architecture of a feed-forward neural network. Draw a neat diagram of a network with one hidden layer (two inputs, two hidden neurons, one output). Briefly explain the role of each component in forward propagation. 8

(c) Compare and contrast the following activation functions commonly used in deep learning :

(i) Sigmoid (logistic) function

(ii) ReLU (Rectified Linear Unit)

(iii) Tanh 8

(d) Consider a single-layer perceptron (no hidden layer) with output

$$\hat{y} = w_1x_1 + w_2x_2 + b$$

and a loss $L = \frac{1}{2}(\hat{y} - y)^2$. Suppose you have one training example where $x_1 = 2, x_2 = s, y = 3$.

(i) Using the standard gradient of squared-error loss, drive $\frac{\partial L}{\partial w_2}$ for this single example as a function of s.

(ii) Plug in $s = 1$ and compute $\frac{\partial L}{\partial w_2}$ numerically. 8

Unit-II

2. (a) Explain how a convolutional layer operates in a CNN. Give a simple diagram showing a 3×3 filter sliding over a 5×5 input. Briefly discuss why convolution layers help in extracting spatial features from images. 4

(b) Describe the role of the pooling layer in a CNN. Explain the two most common type of pooling. 8

(c) Discuss one real-world application of CNNs with the following : What is the input and what is the desired output? Why is a convolutional architecture particularly well-suited for this problem? 8

(d) A grayscale image of size 32×32 is passed through a convolutional layer with the following configuration :

(i) Filter size : 3×3

(ii) Number of filters : 16

(iii) Stride : 1

(iv) Padding : "same" (zero-pad so that output height and width remain 32×32)

(v) Each filter has a bias term (one per filter)

(A) Computer the spatial dimensions (height × width) of each output feature map.

(B) How many parameters (weights + biases) does this convolutional layer have in total?

(C) If the activation after convolution is ReLU, and the input pixel values range [0, 255], what is range of outputs before activation (i.e. the raw convolution sum)? [*Assume all weights and biases lie in [-1, 1]*]. 8

Unit-III

3. (a) Define the architecture of a simple RNN and draw a neat diagram showing how the hidden state is update at each time step. Explain how information flows through time step during forward propagation. 4

(b) Exlain the internal architecture of an LSTM cell. Draw a labelled diagram showing the cell state, hidden state, input gate, forget gate, and output gate. For each gate, write down its activation equation (using sigmoid or tanh where appropriate). 8

(c) Compare a GRU (Gated Recurrent Unit) with an LSTM. Show the GRU update eqautions. List one structural difference (fewer gates, fewer weight matrices) between GRU and LSTM. 8

(d) Describe the structure of the Transformer model as presented in "Attention Is All You Need." 8

Unit-IV

4. (a) Explain the reparameterization trick used in Variational Autoencoder (VAE). Why is it necessary and ho does it enable backpropagation through the sampling operation? 4

(b) Describe the basic architecture and training process of a Autoencoder. Draw a diagram showing encoder, botteneck (latent space) and decoder. Define the reconstruction loss. 8

(c) Discuss the minimax game formulation of a Generative Adversarial Network (GNA). Explain the roles of the Generator and Discriminator. 8

(d) Discuss what mode collapse entails in generative models and outline strategies for preventing it. 8

Unit-V

5. (a) What is hyperparameter tuning in deep learning? Explain why choosing an appropriate learning rate and batch size is critical for successful training. 4

(b) Define L1 and L2 regularization in the context of deep neural networks. Write down the modified loss function for each. Discus how each type of regularization affects the learned weights. 8

(c) Describe the trade-off between model complexity and generalization in deep learning. Explain what is meant by overfitting and underfitting. Illustrate, with

a sketch or description, how training/validation curves behave under the two extremes. 8

(d) A deep neural network has a weight matrix $W \in \mathbb{R}^{3\times 2}$ given by

$$W = \begin{matrix} 2 & -1 \\ 0 & 3 \\ -2 & 4 \end{matrix}$$

Consider adding L2 regularization with coefficient $\lambda = 0.1$ to the weighted sum of squares of all weights. Compute the L2 penalty terms $\frac{\lambda}{2}\sum_{i,j} W_{ij}^2$. 8