JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

MOOC End Term Examination- 2025

B.Tech-VII Semester (CSE/IT)

COURSE CODE(CREDITS):18B2WCI711 MAX. MARKS: 70

COURSE NAME: DEEP LEARNING

COURSE INSTRUCTORS: VANI SHARMA MAX. TIME: 3 Hours

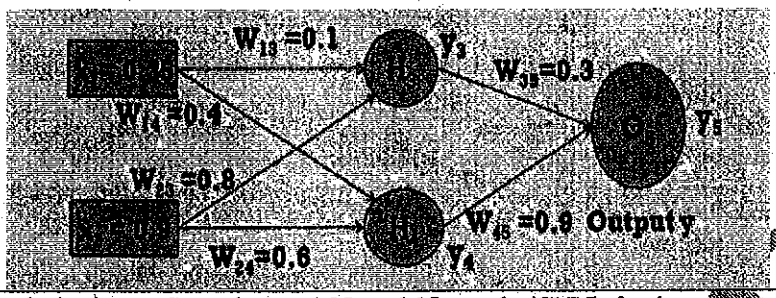
***Note:*** *(a)All questions are compulsory.*

*(b)Marks are indicated against each question in square brackets.*

*(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems*



| Q.No | Question | Marks |
|---|---|---|
| Q1 | (a) Write advantages of momentum based gradient descent, Explain mathematically how momentum based gradient descent is better than the normal gradient descent?<br><br>(b) Give mathematical equations of Adam optimizer, Explain parameters in it, and also explain why bias correction is required? | [3+3] |
| Q2 | (a) How regularization removes the overfitting problem in deep neural network?<br>(b) Discuss Dropout Regularization for deep neural network taking an example neural network.<br>(c) Discuss following algorithms with relation to optimization of training neural network?<br>(i) Mini-Batch algorithm<br>(ii) RMSProp optimization algorithm. | [2+2+4] |
| Q3 | What will happen if we initialize all the weights of a neural network to: (a) Zero, (b) Small random values, (c) Large random values<br>Briefly discuss the effects on the neural network in each case. | [3] |
| Q4 | (a) Show that with a single neuron XOR problem cannot be solved.<br><br>(b) Design a Multi-Layer Perceptron (MLP) neural network for solving XOR function of three inputs X1, X2 and X3 having one hidden layer of four perceptron's and one output layer. Properly show and discuss weights, bias, perceptron's activation functions, etc. | [4+4] |
| Q5 | (a) Compare Sigmoid, Tanh, and ReLU activation functions in terms of their mathematical equations, output range, and gradient behavior.<br><br>(b) What problems arise when using Tanh instead of ReLU in deep neural networks? | [3+2] |
| Q6 | Consider the following ANN model with backpropagation algorithm. Weights and biases are given in the figure. The network uses the sigmoid as an activation function. | [7] |

| | | |
|---|---|---|
| | Use the given information to compute the output of each neuron during the forward pass and then calculate the error term for each neuron using the backpropagation algorithm and update the parameters for one iteration using learning rate ($\eta$) = 0.1.<br>  <br> | |
| Q7 | You are designing a Convolutional Neural Network (CNN) for image classification. The input to the network is an RGB image of size 128 × 128 × 3. The architecture of the first two convolutional layers is given below:<br><br>Layer 1: Number of filters: 16; Filter size: 5 × 5; Stride: 1; Padding: 2<br>Layer 2: Number of filters: 32; Filter size: 3 × 3; Stride: 2; Padding: 1<br><br>(a) Compute the output volume size (height × width × depth) after each convolutional layer. Show all intermediate steps and formulas used.<br>(b) Compute the total number of learnable parameters (including biases) for both layers combined.<br>(c) If a 2×2 max pooling layer (stride = 2) is applied after Layer 2, what will be the new output volume dimensions?<br>(d) Briefly explain how changing the stride and padding in Layer 2 would affect the spatial resolution and computational cost of the network. | [8] |
| Q8 | (a) Describe the various possible architectures for input and output combinations of an RNN for non-sequence and sequence types. Discuss sequence to sequence models for same length and different length input and outputs. Give applications of each architecture.<br><br>(b) Discuss how you would train the recurrent neural network for sentiment classification on Amazon product reviews.<br><br>(c) Explain how a Deep RNN differs from a Simple RNN in terms of architecture, feature representation, and training difficulty. | [5+3+3] |
| Q9 | (a) Describe the architecture of LSTM with a neat and clean diagram.<br>(b) Discuss the significance of all the gates used along with mathematical equations.<br>(c) Does the LSTM solve the issue of poor long-term memory in RNNs? | [2+3+2] |
| Q10 | Describe the concept of cross attention in the transformer architecture. How does masked multi-head attention prevent information leakage during model training? | [7] |