JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION- 2024

M.Tech-IISemester (CSE/IT/~~ECE/CE~~) BS

COURSE CODE(CREDITS): 22M11CI212 MAX. MARKS: 25

COURSE NAME: Deep Learning Techniques

COURSE INSTRUCTORS: HRI MAX. TIME: 1 Hour 30 Minutes

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

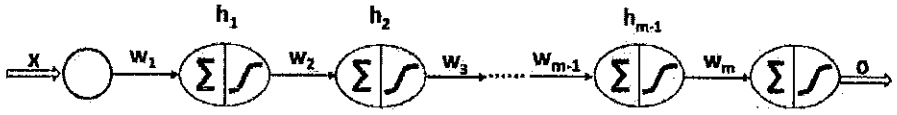
(b)Marks are indicated against each question in square brackets.

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

---

Q1. In a MLP network shown below, show that **[CO2][04 Marks]**

$$\frac{\partial L}{\partial h_t} = \Phi'(h_{t+1}) \cdot w_{t+1} \cdot \frac{\partial L}{\partial h_{t+1}}$$



Q2. Show with a mathematical explanation that the sigmoid and tanh activation functions suffer from the vanishing gradient problem. **[CO2][04 Marks]**

Q3. Write advantages of momentum based gradient descent. Explain mathematically as how is momentum based gradient descent better than the normal gradient descent?**[CO2][2+3=05 Marks]**

Q4. Suppose an input image has been converted into a matrix of size 256 x 256 and a kernel/filter of size 3X3 with a stride of 1 and padding of 0 is convolved over this image. What will be the size of the convoluted matrix? **[CO3][04 Marks]**

Q5. A convolution operation is performed over an input gray scale image of size 6x6 (represented as matrix X) with a filter of size 3x3 representing its weight matrix w1 and bias b1 that results in the next layer feature map z1. Then ReLU activation function is applied over the z1 and the output activation A1 is obtained. Then Maxpooling operation over the A1 with a filter of 2x2 results in the feature map P1. Then P1 is flatten and a 1-Dimensional vector F is obtained. Then the features in the F are connected to a single perceptron having bias b2 and weights over all these connections are represented with a 1-Dimensional matrix w2. It results in the pre-activation z2. At last the sigmoid activation function is applied to make a binary classification and the loss is computed as the binary cross entropy.

Write a mini-batch gradient descent solution to update all trainable parameters in the above CNN architecture. **[CO3][08 Marks]**