JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
TEST -2 EXAMINATION – October 2023
B.Tech-VII Semester Elective (CSE/IT/ECE/CE/BT/BI)

COURSE CODE / (CREDITS): 19B1WCI738 / (3) MAX. MARKS: 25
COURSE NAME: Introduction to Deep Learning
COURSE INSTRUCTORS: HRI, KLK, VKS 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. Describe the following about a ReLU function. **(CO1) [03 Marks]**
(a) Mathematical expression (b) domain (c) range (d) graph (e) derivative and its graph.

Q2. Derive expression for the weight update rule for binary classification via logistic regression implanted through perceptron. Given n pairs of (X_i, y_i), where X_i is a d-dimensional feature vector and y_i is binary target variable with values +1 and -1. **(CO1) [03 Marks]**

Q3. Show with a mathematical explanation as how the activation functions which are not mean centric have slow convergence. **(CO2) [03 Marks]**

Q4. Describe with a mathematical explanation the Adadelta Gradient Descent. **(CO2) [03 Marks]**

Q5(a). An input image has been converted into a matrix of size 224 x 224 and convolved with a kernel of size FxF with a stride of S and padding of P to produce a feature map of dimension 222 x 222. Find out the values of filter size, stride rate and padding? **(CO3) [03 Marks]**

Q5(b). Suppose in a layer you have an input volume of size 32 x 32 x 3, ten 5x5 kernels/filters with a stride of 1 and padding of 2. Find out the number of parameters in this layer? **(CO3) [03 Marks]**

Q6 (a). Is CNN better than ANN in terms of weight sharing and sparsity. The sparsity means each output in a layer comes from a small number of inputs? **(CO3) [03 Marks]**

Q6(b). A convolution operation is performed over an input gray scale image of size 3 x 3 (represented as matrix X) with a filter of size 2 x 2 representing its weight matrix w1 and bias b1 that results in the next layer feature map z_1. Then after the ReLU, Maxpooling and flatten the 1-Dimensional flatten vector is fed to a single perceptron. At last the sigmoid activation function is applied to make a binary classification and the loss (L) is computed as the binary cross entropy.
Assume that during the back propagation the derivative of loss with respect to z_1 is known or already computed and is as follows.

$$\frac{\partial L}{\partial z_1} = \begin{bmatrix} \frac{\partial L}{\partial z_{11}} & \frac{\partial L}{\partial z_{12}} \\ \frac{\partial L}{\partial z_{21}} & \frac{\partial L}{\partial z_{22}} \end{bmatrix}$$

Write a mini-batch back propagation gradient descent solution to update the w1 and b1 trainable parameters in the above CNN architecture. **(CO3) [04 Marks]**