National University of Computer & Emerging Sciences, Karachi
Spring-2023 School of Computing
Mid I Examination
27th February 2023, 08:30 am – 09:30 am

Course Code: CS4045	**Course Name:** Deep Learning
Instructor Name: Dr. Muhammad Atif Tahir, Ms. Sumaiyah Zahid	
Student Roll No:	**Section:**

Instructions:

- Return the question paper and make sure to keep it inside your answer sheet.
- Read each question completely before answering it. There are a total of **three questions on two pages**.
- In case of any ambiguity, you may make assumptions. However, your assumption should not contradict any statement in the question paper.
- Do not write anything on the question paper (except your ID and section). You will be graded ONLY on the answer sheet.

Total Time: 1 Hour **Max Points**: 10 Marks

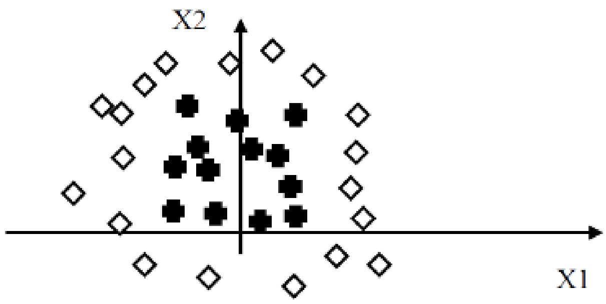
Question # 1 **[CLO 1](2 marks)**

Briefly answer the following questions. Each question should be answered in 3 – 4 lines including articles. Otherwise, answers will not be checked.

A. Briefly discuss why it is possible to design logistic regression classifiers using neural network algorithms. For example, spam detection can be done using logistic regression. But also it can be done using a neural network, so how this is possible?
Ans: Neural Network is a universal approximator so it is possible to implement logistic regression using neural network
B. Explain the difference b/w discriminative and generative classifiers. Which category does logistic regression belong to?
Ans: A generative model like naive Bayes makes use of this likelihood term, which expresses how to generate the features of a document if we knew it was of class c. By contrast a discriminative model in this text categorization scenario attempts to directly compute P(c|d). LR = discriminative
C. Briefly discuss the training and testing phase of Logistic Regression.
training: we train the system (specifically the weights w and b) using stochastic gradient descent and the cross-entropy loss.
test: Given a test example x we compute p(y|x) and return the higher probability label y = 1 or y = 0.
D. Is it possible to use ReLU activation in the last layer for (i) classification and (ii) Clustering?
For unsupervised learning, yes. For classification NO

Question # 2 **[CLO 1](2 marks)**

Suppose that we want to build a neural network that classifies two-dimensional data (i.e., X = [x1, x2]) into two classes: diamonds and crosses. We have a set of training data that is plotted as follows:



Draw a network that can solve this classification problem. Justify your choice of the number of nodes and the architecture.

Solution: A solution is a multilayer FFNN with 2 inputs, one hidden layer with 4 neurons and 1 output layer with 1 neuron. The network should be fully connected, that is there should be connections between all nodes in one layer with all the nodes in the previous (and next) layer. We have to use two inputs because the input data is two dimensional. We use an output layer with one neuron because we have 2 classes. One hidden layer is enough because there is a single compact region that contains the data from the crosses-class and does not contain data from the diamonds-class. This region can have 4 lines as borders, therefore it suffices if there are 4 neurons at the hidden layer. The 4 neurons in the hidden layer describe 4 separating lines and the neuron at the output layer describes the square that is contained between these 4 lines.

Question # 3 **[CLO 2](6 marks)**

Consider the following neural network with 2 hidden layers. The relu activation function is used in hidden layers 1 and 2. Sigmoid activation is used in the output layer. The details of the input and weights are given below. Biases are 0 for all neurons.
X = [[0.55], [0.72],[0.93]]
weights of L1
W1 = [[0.42, 0.72, 0.21], [0.11, 0.3, 0.65], [0.15, 0.09, 0.36], [0.32, 0.51, 0.12]]
weights of L2
W2 = [[0.17, 0.29, 0.35, 0.51], [0.91, 0.08, 0.51, 0.37], [0.61, 0.39, 0.16, 0.01], [0.38, 0.21, 0.82, 0.14]]
weights of L3
W3 = [[0.44, 0.03, 0.55, 0.81]]
Learning rate = 0.7

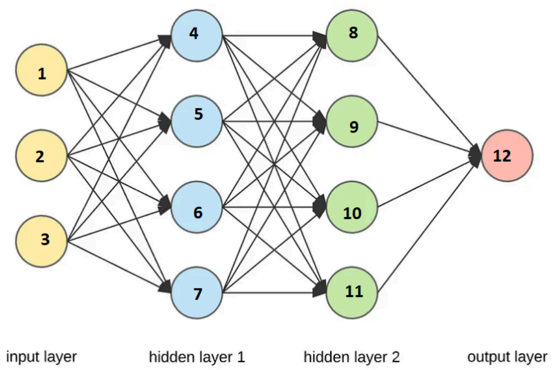
Sigmoid activation function and it's derivative is:

$\sigma(x) = \frac{1}{1+e^{-x}}$ [1]	$f(x)(1-f(x))$

Relu activation function and it's derivative is:

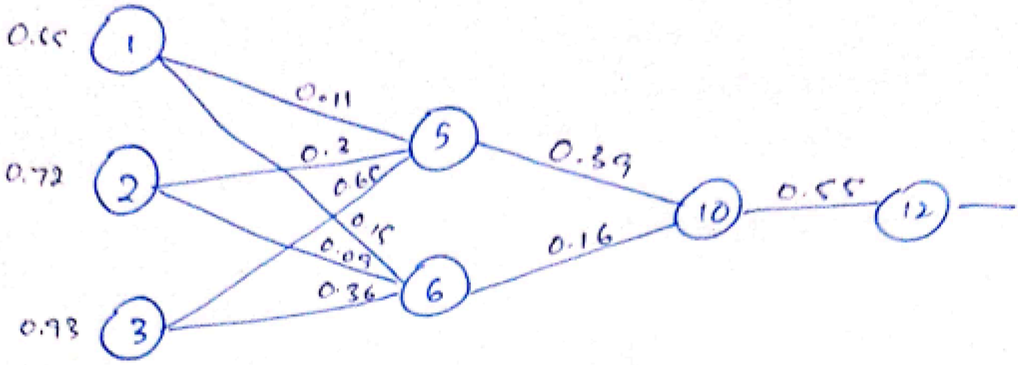
$\begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} = \max\{0, x\} = x\mathbf{1}_{x>0}$	$\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \\ \text{undefined} & \text{if } x = 0 \end{cases}$

Loss function and it's derivative is:
Loss= ½(y-a)^2
dLoss = - (y - a)



Use the dropout probability for 0.5 for hidden layer 1 and 0.75 for hidden layer 2. All forward and backward propagation formulas should be shown clearly.
i. [1 Point] Compute forward propagation on the thinned network.
ii. [1 Point] Compute the error using the squared error function if y=0.98.
iii.[3 Points] Perform backpropagation on a thinned network and update the weights.

iv.[1 Point] Test the updated network on new test data if X=[[0.33], [0.45], [0.81]]. Compute the output.



$z_5 = w_{15} \times x_1 + w_{25} x_2 + w_{35} x_3 \Rightarrow 0.881$

$a_5 = 0.881$

$z_6 = w_{16} x_1 + w_{26} x_2 + w_{36} x_3 \Rightarrow 0.4821$

$a_6 = 0.4821$

$z_{10} = w_{5\,10}\, a_5 + w_{6\,10}\, a_6 \Rightarrow 0.420726$

$a_{10} = 0.420726$

$z_{12} = 0.420726 \times w_{10\,12}$

$z_{12} = 0.2313993$

$$a_{12} = \frac{1}{1+e^{-0.2313993}} \Rightarrow \boxed{0.55759}$$

Part a (i

$$\mathcal{E} = \frac{1}{2}\left[0.78 - 0.557597\right]^2$$

$$\boxed{\mathcal{E} = 0.089213}$$

part (ii)

Backpropagation

$$\frac{\partial L}{\partial w_{1012}} = -(y - a_{12})\, a_{12}(1 - a_{12})\, a_{10}$$

$$\frac{\partial L}{\partial w_{1012}} = -0.04384$$

$$w_{1012}^{+} = w_{1012} - \eta \frac{\partial L}{\partial w_{1012}}$$

$$= 0.55 - (0.7)(-0.04384)$$

$$\boxed{w_{1012} = 0.581}$$

$$\frac{\partial L}{\partial b_{12}} = -(y - a_{12})\, a_{12}(1 - a_{12})$$

$$\frac{\partial L}{\partial b_{12}} = -0.1042$$

$$b_{12} = 0 - (0.7)(-0.1042)$$

$$\boxed{b_{12} = 0.07294}$$

derivative of relu

$$\frac{\partial L}{\partial w_{510}} = -(y-a_{12})a_{12}(1-a_{12})\, w_{1012} \times 1 \times a_5$$

$$= -0.0505$$

$$\boxed{w_{510}^{+} = 0.42535}$$

$$\frac{\partial L}{\partial w_{610}} = -(y-a_{12})a_{12}(1-a_{12})\, w_{1012} \times 1 \times a_6$$

$$= -0.0276$$

$$\boxed{w_{610}^{+} = 0.17932}$$

$$\frac{\partial L}{\partial b_{10}} = -(y-a_{12})a_{12}(1-a_{12})\, w_{1012} \times 1$$

$$= -0.0573$$

$$\boxed{b_{10}^{+} = 0.0401}$$

$$\frac{\partial L}{\partial w_{15}} = -(y-a_{12})a_{12}(1-a_{12})\, w_{1012} \times 1 \times w_{510} \times 1 \times x_1$$

$$= -0.01229$$

$$\boxed{w_{15}^{+} = 0.118603}$$

$$\frac{\partial L}{\partial w_{25}} = -(y-a_{12})a_{12}(1-a_{12})\, w_{1012} \times 1 \times w_{510} \times 1 \times x_2$$

$$= -0.0161$$

$$\boxed{w_{25}^{+} = 0.31127}$$

$$\frac{\partial L}{\partial w_{35}} = -(y-a_{12})a_{12}(1-a_{12})w_{1012} \times 1 \times w_{510} \times 1 \times x_3$$

$$= -0.02078$$

$$\boxed{w_{35}^{+} = 0.664546}$$

$$\frac{\partial L}{\partial b_5} = -(y-a_{12})a_{12}(1-a_{12})w_{1012} \times 1 \times w_{510} \times 1$$

$$= -0.02235$$

$$\boxed{b_5^{+} = 0.015645}$$

$$\frac{\partial L}{\partial w_{16}} = -(y-a_{12})a_{12}(1-a_{12})w_{1012} \times 1 \times w_{610} \times 1 \times x_1$$

$$= -0.00504$$

$$\boxed{w_{16}^{+} = 0.153528}$$

$$\frac{\partial L}{\partial w_{26}} = -(y-a_{12})a_{12}(1-a_{12})w_{1012} \times 1 \times w_{610} \times 1 \times x_2$$

$$= -0.0066$$

$$\boxed{w_{26}^{+} = 0.09462}$$

$$\frac{\partial L}{\partial w_{36}} = -(y-a_{12})a_{12}(1-a_{12})w_{1012} \times 1 \times w_{610} \times 1 \times x_3$$

$$= -0.008527$$

$$\boxed{w_{36}^{+} = 0.36596}$$

$$\frac{\partial L}{\partial b_6} = -(y-a_{12})a_{12}(1-a_{12})\, w1012 \times 1 \times w610 \times 1$$

$$= -0.00916$$

$$\boxed{b_6^+ = 0.006412}$$

iv) Testing

Updated weights, multiplying by P

W1 = [[0.21, 0.36, 0.105], [0.0593, 0.15563, 0.33273]
[0.076764, 0.04731, 0.18278], [0.16, 0.255, 0.06]]

w2 = [[0.1275, 0.2175, 0.2625, 0.3825],
[0.6825, 0.06, 0.3825, 0.2775],
[0.4575, 0.3190125, 0.13449, 0.0075],
[0.285, 0.1575, 0.615, 0.105]]

w3 = [0.44, 0.03, 0.581, 0.81]

$z_4 = 0.31635$

$a_4 = 0.31635$

$z_5 = 0.3747588$

$a_5 = 0.3747588$

$z_6 = 0.20124742$

$a_6 = 0.20124742$

$z_7 = 0.21615$

$a_7 = 0.21615$

$z_8 = 0.25734$

$a_8 = 0.25734$

$z_9 = 0.37535$

$a_9 = 0.37535$

$z_{10} = 0.293$

$a_{10} = 0.293$

$z_{11} = 0.295647$

$a_{11} = 0.295647$

$z_{12} = 0.60713$

$\boxed{a_{12} = 0.64728}$

Be like a neural network. Learn from your mistakes.
***** Best of Luck *****