National University of Computer and Emerging Sciences

Deep Learning For Perception (CS4045)

Date: February 28th 2024

Course Instructor(s)

Ms. Sumaiyah Zahid

Sessional-I Exam

Total Time: 1 Hour
Total Marks: 15
Total Questions: **03**

Semester: SP-2024
Campus: Karachi
Dept: Computer Science

_____	_____	_____	_____
Student Name	Roll No	Section	Student Signature

_____	_____
Vetted by	Vetter Signature

CLO # 1: Student should be able to describe what Deep Learning is and the skill sets needed for Deep Learning

Q1: Briefly answer the following questions. Each question should be answered in 3 – 4 lines including articles. Otherwise, answers will not be checked. **[0.5*6=3 marks]**

1. Is it true that the number of neurons in the output layer should match the number of classes i.e. where is the number of classes are greater than 2.
2. What are hyper-parameters? Name any three.
3. Is it possible to use ReLU in the last layer?
4. Write a derivative equation for log loss with L2 regularization.
5. What Is the Difference Between Epoch, Batch, and Iteration in Deep Learning?
6. List two drawbacks of DropConnect.

CLO # 2: Students should be able to understand supervised and unsupervised methods of Deep Learning

Q2: **[1+1+1 =3 marks]**

A standard feed-forward neural network is trained by Mini Batch Gradient Descent. The network has 9 input neurons, 5 output neurons and a hidden layer, H. Hidden layer use relu activation while output layer uses softmax activation function. The network is fully connected, but only the hidden layer uses biases. The network has a total of 300 trainable parameters, pi ∈ P. During a particular session, it

performs the forward and backward processes of on a small batch size of 40 samples out of total 200 observations.

1. How many nodes are in hidden layer H?
2. How many total internal calculations of gradients of the form ∂L/ ∂pi are performed in one mini batch where pi is any trainable parameter?
3. How many total parameter updates (of the pi) are performed in a single mini batch?

CLO # 2: Students should be able to understand supervised and unsupervised methods of Deep Learning

Q3: **[1+1+4+3=9 marks]**

Consider the following feed forward neural network:
Consider the given two-layer architecture, where each node in the layer L1 applies ReLU non linearity and L2 uses the Hyperbolic Tangent activation function, for simplicity the bias unit is not considered for the given network.

X = [[0.55], [0.72],[0.81]]

weights of L1

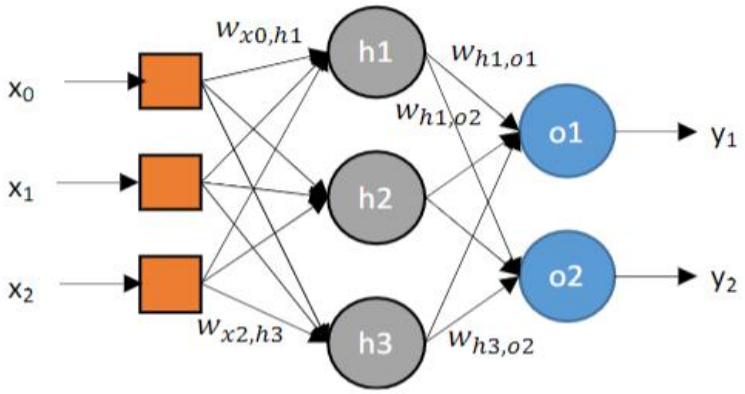
W1 = [[0.42, 0.72,0.5],[0. , 0.3 , -0.2],
[-0.34,0.15, -0.9]]

weights of L2

W2 = [[-0.97, 0.55, 0.97],[0.71, -0.7, 0.22]]

Learning Rate = 0.5.

Regularization Coefficient= 0.03



Hyperbolic Tangent activation and its deriv

$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$1 - f(x)^2$

Relu activation and its derivative is:

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

The squared error loss is given below

$$J(\theta) = \frac{1}{2m} \sum \left(y_i - \mathrm{a}_i^{(2)} \right)^2$$

1. Compute the output of each node in the network for forward propagation.
2. Compute the error if y= [-0.98, 0.75]
3. Perform backpropagation, and compute the derivatives i.e. dWh1,o1, dWh2,o2, dWx0,h1. Calculate the updated weights.
4. Perform backpropagation with l1 regularization and compute the derivatives for dWh2,o1, dWh3,o2. Calculate the updated weights.