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

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.
   Ans: False: Depends upon output coding. For 4 class problem, two output neurons are enough
2. What are hyper-parameters? Name any three.
   Ans: Hyperparameters are configuration settings in machine learning models. Three examples are: Learning Rate, Number of Hidden Layers and Neurons, Batch Size
3. Is it possible to use ReLU in the last layer?
   Ans: For unsupervised learning, yes. For classification NO
4. Write a derivative equation for log loss with L2 regularization.
   Ans: (a-y)/(a(1-a) + (-) lambda w
5. What Is the Difference Between Epoch, Batch, and Iteration in Deep Learning?
   Ans: Epoch - Represents one iteration over the entire dataset (everything put into the training model). Batch - Refers to when we cannot pass the entire dataset into the neural network at once, so we divide the dataset into several batches. Iteration - if we have 10,000 images as data and a batch size of 200. Then an epoch should run 50 iterations (10,000 divided by 50)
6. List two drawbacks of DropConnect.
   Ans: DropConnect's drawbacks: Training with DropConnect is slower. DropConnect's implementation is more complicated than Dropout's implementation. DropConnect has been proven to work mostly when using more than one network.

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?

Solution:

9H + H + 5H = 300, so H = 20

300 x 40 = 12,000

300, each pi one time

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

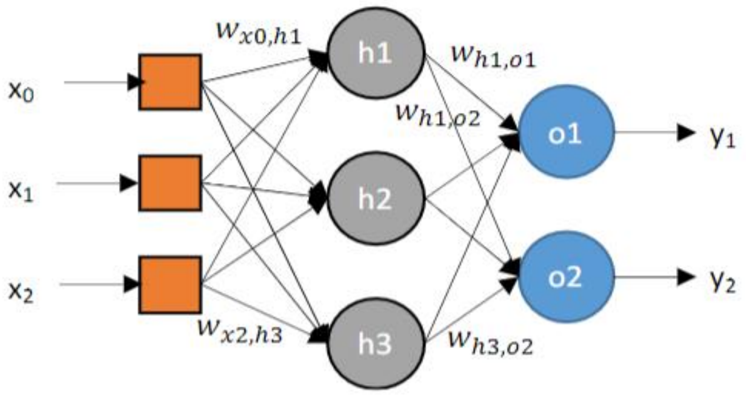
W1 = [[0.42, 0.72,0.5],[0. , 0.3 , -0.2],
[-0.34,0.15, -0.9]]

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 derivative is:

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

Relu activation and its derivative is:

$$f(x) = \begin{cases} 0 \text{ for } x<0 \\ x \text{ for } x \geq 0 \end{cases}$$

$$f'(x) = \begin{cases} 0 \text{ for } x<0 \\ 1 \text{ for } x \geq 0 \end{cases}$$

The squared error loss is given below

$$\text{MSE} = \frac{1}{2N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

1. Compute the output of each node in the network for forward propagation.
   Z(h1)=1.1544→ a(h1)=1.1544
   Z(h2)=0.116→ a(h2)=0.116
   Z(h3)=-0.4848→ a(h2)=0
   Z(o1)=-1.0559→ a(o1)=-0.784
   Z(o2)=-0.7384→ a(o2)=0.63819
2. Compute the error if y= [-0.98, 0.75]
   E= ½ ((-0.784+0.98)^2 + (0.63819-0.75)^2)=0.02545
3. Perform backpropagation, and compute the derivatives i.e. dWh1O1, dWh2O2, dWx0h1. Calculate the updated weights.
   D h1o1=-(-0.98+0.784)(1-0.784^2) 1.1544 = 0.0871
   Wh1o1= -0.97 - 0.5*0.0871=-1.01355
   D h2o2=-(0.75-0.63819)(1-0.63819^2) 0.116= -0.00768
   Wh2o2= -0.7- 0.5*-0.00768=-0.69616
   D X0h1=-(-0.98+0.784)(1-0.784^2) (-0.97)(1)(0.55) + (-(0.75-0.63819)(1-0.63819^2) (0.71)(1)(0.55))
   D X0h1=-0.0662
   W X0h1= 0.42- 0.5*-0.0662=0.4531
4. Perform backpropagation with **L1 regularization** and compute the derivatives for dWh2O1, dWh3O2. Calculate the updated weights.
   D h2o1=-(-0.98+0.784)(1-0.784^2) 0.116 + 0.03 = 0.03876
   Wh2o1= 0. - 0.5*0.03876=-0.01938
   D h3o2=0 + 0.03 = 0.03
   Wh3o2= 0.22 - 0.5*0.03=0.205