National University of Computer and Emerging Sciences

Deep Learning For Perception (CS4045)

Date: April 1st 2024

Course Instructor(s)

Ms. Sumaiyah Zahid

Sessional-II Exam

Total Time: 1 Hour
Total Marks: 10
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: Write short answers in maximum 3 to 4 lines. [0.5*6=3 marks]

1. Discuss the use of residual connections in CNN architectures. How do they help address the vanishing gradient problem?

Sol: Residual connections in CNN architectures facilitate the flow of gradients during training by creating shortcut connections that bypass certain layers. This helps mitigate the vanishing gradient problem by providing an alternative path for gradient flow, enabling more efficient optimization of deep networks.

2. What is the concept of inception block in Googlenet? How 1*1 convolution helps in reducing the complexity of network?

Sol: In GoogLeNet's Inception blocks, 1x1 convolutions are used for dimensionality reduction, effectively reducing network complexity by decreasing parameters while allowing for multi-scale feature blending, thereby enhancing representational capacity.

3. What must be the value of forget gate and input gate if we want to retain any information indefinitely?

Sol: The forget gate value should be set to 1, meaning forget nothing from the previous cell state, and the input gate value should be set to 0.

4. In an LSTM network if W has the dimension of 4*3 and U has the dimension of 4*2 and biases have a dimension of 4*1. What would be the dimension of x_t and out_{t-1}?

$\partial x_t = W^T . \partial gate\ s_t$

$\Delta out_{t-1} = U^T . \partial gate\ s_t$

Sol: xt = 3*1
out t-1=2*1

5. Write three applications of many to many LSTM/RNN network.

Sol: Machine Translation, Video Activity Recognition, Speech Recognition

6. Why bidirectional RNNs are preferred over forward RNN in sentiment analysis problem?

Sol: To learn the context from both sides.

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

Q2: **[1+1+2 = 4 marks]**

1. Consider a CNN architecture, where the input image is of size 200x200x64. The first convolutional layer has 96 kernels of size 10x10x3. The stride is 2 and padding is 1. What would be the size of the output image after the first convolutional layer?
   Sol: output = (input + 2p –k)/s+1
   output = (200 + 2(1) – 10) / 2 +1 = 192/2+1 = 97x97x96
2. Calculate the number of parameters with RGB image as input, followed by 3×3 filter, and output with 9 channels. **Num_params= [i* (f*f)*o] +o**
   Sol: num_params
   = [i × (f×f) × o] + o
   = [3 × (3×3) × 9] + 9
   = 252

3. Consider simple convolutional layer example with following input X and Filter F

X= [[2 3 1]
[4 5 6]
[4 7 8]]

F= [[1 -1]
[-1 1]]

Find the gradients for X and F i.e. ∂L/∂X and ∂L/∂F assuming ∂L/∂O = [[2 3],[4 5]].

$$\frac{\partial L}{\partial F} = \text{Convolution}\left(\text{Input } X\text{ , Loss gradient } \frac{\partial L}{\partial O}\right)$$

$$\frac{\partial L}{\partial X} = \text{Full Convolution}\left(\text{180° rotated Filter } F\text{ , Loss Gradient } \frac{\partial L}{\partial O}\right)$$

Sol: ∂L/∂X = [[2,1,-3],[2,0,-2],[-4, -1,5]]
∂L/∂F=[[54,59],[74,96]]

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

Q3: **[2+1=3 marks]**

Imagine we discover aliens speaking a language that's the opposite of ours! To figure out their feelings when they talk, we're using a special system called a backward RNN. Our job is to show how this system works step by step with the alien talk. We're trying to find out what emotions the aliens express in their reversed language. (positive, neutral or negative)

Here's a sample of their talk: "ereht olleH" // Let's assume this sentence is neutral
For simplicity, biases are considered 0.

W_xh = [[0.4, 0.1],
[-0.2, 0.3],
[0.5, 0.1],
[0.4, -0.3]]

W_hy = [[0.2, 0.6, -0.1,0.2],
[0.3, -0.2, 0.4,0.5],
[0.1, 0.2, 0.3,0.9]]

W_hh = [[0.2, -0.1, 0.3,0.2],
[-0.1, 0.4, -0.2,-0.3],
[0.5, 0.1, 0.7,0.3],
[0.4, -0.3, 0.5,0.1]]

h_0 = [0.1, -0.1, 0.2,0.5]

1. Show all computations for RNN output for 2 timestamps.
2. Calculate the final error.

h_t = tanh(W_xh * x_t + W_hh * h_(t-1))
pos_scores = W_hy * h_t
Cross-Entropy Loss (L) = -Σ(T_i * log(P_i))

Sol:
One Hot Encoding :
There=[1,0]
Hello = [0,1]

Hello will be passed first because its backward RNN.

H_1= tanh ([[0.4, 0.1],
[-0.2, 0.3],
[0.5, 0.1],
[0.4, -0.3]] * [0,1] + [[0.2, -0.1, 0.3,0.2],
[-0.1, 0.4, -0.2,-0.3],
[0.5, 0.1, 0.7,0.3],
[0.4, -0.3, 0.5,0.1]] * [0.1, -0.1, 0.2,0.5])

H_1= tanh ([0.1,0.3, 0.1 , -0.3] + [0.19 ,-0.24, 0.33 , 0.22])
H_1= [0.282, 0.059, 0.405, -0.0798]

H_2= tanh ([[0.4, 0.1],
[-0.2, 0.3],
[0.5, 0.1],
[0.4, -0.3]] * [1,0] + [[0.2, -0.1, 0.3,0.2],
[-0.1, 0.4, -0.2,-0.3],

[0.5, 0.1, 0.7,0.3],
[0.4, -0.3, 0.5,0.1]] * [0.282, 0.059, 0.405, -0.0798])

H_2= tanh ([0.4, -0.2, 0.5 , 0.4] + [0.156, -0.0616 , 0.4065, 0.28962])
H_2= [0.505, -0.2558, 0.7194, 0.5973]

pos_scores = [[0.2, 0.6, -0.1,0.2],
[0.3, -0.2, 0.4,0.5],
[0.1, 0.2, 0.3,0.9]] * [0.505, -0.2558, 0.7194, 0.5973]

pos_scores =[-0.00496, 0.78907, 0.75273]
Softmax =[0.187 , 0.4138 ,0.3991]

Cross-Entropy Loss (L) = -Σ(T_i * log(P_i))
L = -(0 * log(0.187) + 1 * log(0.4138) + 0 * log(0.3391))
L = -(log(0.4138))
L ≈ 0.383 or 0.882 with ln