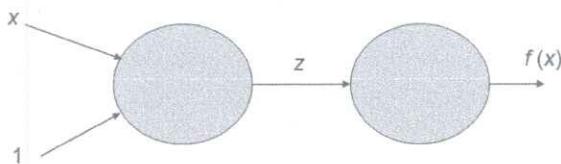


**The LNM Institute of Information Technology, Jaipur**  
**CSE4121: Deep Learning**  
**Mid term Exam**

**Max Marks: 20****Date: 10/03/2023****Time: 90 Mins**

Total Questions	Total Marks	CO1	CO2	CO3	CO4
5	20	Q1, Q2	Q3, Q4, Q5	--	--
<b>CO weightage</b>	<b>12/20 = 60%</b>		<b>8/20 = 40%</b>		

Q1. Consider the following computation:



Here,  $z = (1 + \tanh(wx + b))/2$  and  $f(x) = \text{sig}(z)$ . By definition  $\text{sig}(z) = 1/(1 + e^{-z})$  and  $\tanh(z) = (e^z - e^{-z})/(e^z + e^{-z})$ . Value  $L$  is given by  $L = -y \log(f(x))$ . Here,  $x$  and  $y$  are constants and  $w$  and  $b$  are parameters that can be modified. Derive the partial derivative  $\partial L/\partial w$  and  $\partial L/\partial b$ . (8)

Q2. Answer the following briefly: (2\*2 = 4)

- a) You are training a large feedforward neural network (100 layers) on a binary classification task, using a *sigmoid* activation in the final layer, and a mixture of *tanh* and *ReLU* activations for all other layers. You notice the weights to a subset of layers stop updating after the first epoch of training, even though the network has not yet converged. Deeper analysis reveals the gradients to these layers completely, or almost completely, go to zero very early on in training. What is the problem and how can you fix it?
- b) You are working on an automated check-out system for a supermarket, and are building a classifier for apples, bananas and oranges. Suppose your classifier obtains a training set error of 0.5%, and a dev set error of 7%. What can you do to try to improve your classifier?

Q3. Suppose the temperature in Casablanca over the first three days of January are the same:

Jan 1<sup>st</sup>:  $\theta_1=10^\circ C$ , Jan 2<sup>nd</sup>:  $\theta_2=10^\circ C$

You use an exponentially weighted average,  $v_t$ , with  $\beta=0.9$  to track the temperature with  $v_0=0$ . If  $v_2$  is the value computed at end of day 2 what will be its value? (2)

Q4. Explain the bias-variance trade-off using the graph of Error vs. Model complexity. (3)

Q5. Give brief answers for the following: (1\*3 = 3)

- a) Why is dropout used in deep learning?
- b) When and why is Leaky ReLU function used?
- c) How does Batch normalization help in the learning process?