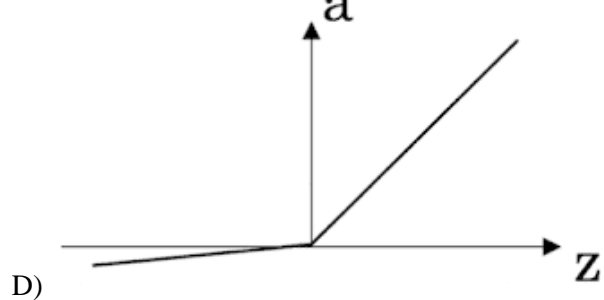
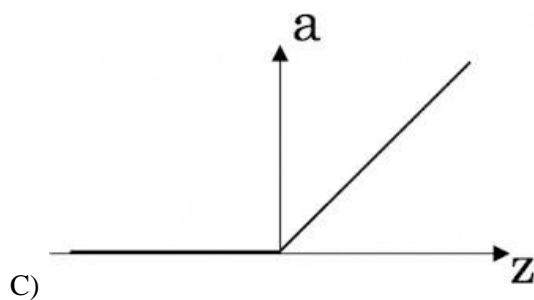
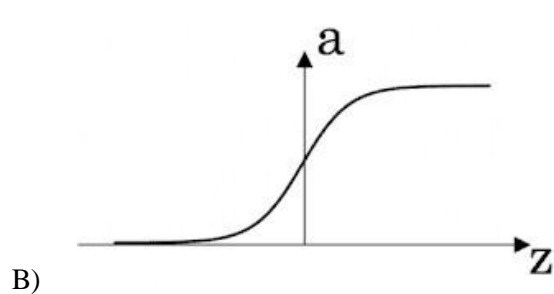
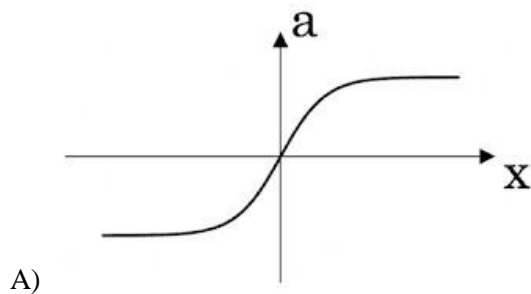


# WORKSHEET-2

## DEEP LEARNING

**Q1 to Q8 are MCQs with only one correct answer. Choose the correct option.**

1. Operations in the neural networks can performed\_\_\_\_\_?  
 A) serially  
 B) parallely  
 C) serially or parallely  
 D) None of the above  
 Answer: C) serially or parallely
2. Who proposed the first perceptron model and when?  
 A) Rosenblatt, 1958  
 B) McCulloch-pitts, 1958  
 C) John Hopfield, 1982  
 D) McCulloch-pitts, 1982  
 Answer:A)
3. Which one of these plots represents a ReLU activation function?



Answer: C)

4. In a simple artificial neural network with 5 neurons in the input layer, 8 neurons in the hidden layer and 3 neurons in the output layer. What is the size of the weight matrices between hidden-output layers and input-hidden layers?  
 A)  $[3 \times 8]$ ,  $[5 \times 8]$   
 B)  $[8 \times 3]$ ,  $[5 \times 8]$   
 C)  $[5 \times 8]$ ,  $[8 \times 5]$   
 D)  $[8 \times 3]$ ,  $[5 \times 3]$   
 Answer D)  $[8 \times 3]$ ,  $[5 \times 3]$
5. What is a dead unit in a neural network?  
 A) A unit which does not respond completely to any of the training patterns  
 B) The unit which produces the biggest sum-squared error  
 C) A unit which doesn't update during training by any of its neighbour  
 D) None of these  
 Answer:A)
6. Which of the following functions can be used as an activation function if we wish to predict the probabilities of n classes such that sum of all n probabilities is equal to 1?  
 A) sigmoid  
 B) softmax  
 C) tanh  
 D) ReLU  
 Answer: B)
7. The amount of output of one unit received by another unit depends on what?  
 A) output unit  
 B) input unit  
 C) activation values  
 D) weights  
 Answer: D)

8. What is asynchronous update in neural networks?

- A) output units are updated parallely
- B) output units are updated sequentially
- C) either sequentially or parallely
- D) None of the above

Answer:B)

**Q9 and Q10 are MCQs with one or more correct answers. Choose all the correct options.**

9. Which of the following techniques can be used to reduce overfitting in a neural network?

- A) EarlyStopping
- B) Dropout
- C) checkpoints
- D) ReduceLROnPlateau

Answer: A) & B)

10. Why is an RNN used for machine translation, say translating English to Hindi?
- A) It can be trained as a supervised learning problem.
  - B) It is strictly more powerful than a Convolutional Neural Network
  - C) It is applicable when the input/output is a sequence (e.g., a sequence of words)
  - D) RNNs represent the recurrent process of Idea->Code->Experiment->Idea->....
- Answer: A) & C)

**Q11 to Q15 are subjective answer type question. Answer them briefly.**

11. The output of a perceptron is calculated as follows:

$$y = f\left(b + \sum_{i=1}^n w_i x_i\right)$$

Where  $f(x)$  is the activation function. If you want to build a perceptron which gives an output for linear

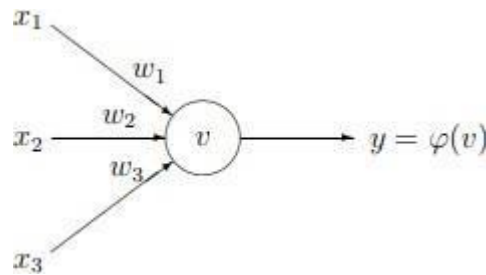
regression, what will be the activation function you would use?

**Answer:** Linear Activation Function

12. What will happen if we use very large or very small learning rates?

**Answer:** The learning rates are part of optimization process smaller learning rates requires. Conversely, larger learning rates will require fewer training epochs. Further, smaller batch sizes are better suited to smaller learning rates given the noisy estimate of the error gradient. A traditional default value for the learning rate is 0.1 or 0.01, and this may represent a good starting point on your problem.

13. Below is a diagram of a single artificial neuron:



The node has three inputs  $x = (x_1, x_2, x_3)$  that receive only binary signals (either 0 or 1). How many different input patterns this node can receive? What if the node had four, five inputs? Can you give a formula that computes the number of binary input patterns for a given number of inputs?

**Answer:** Nodes can have 3 inputs or even more  $y = a(b + w_1x_1 + w_2x_2 + w_3x_3)$

14. What Are Vanishing and Exploding Gradients?

**Answer: Exploding Gradient** An error gradient is the direction and magnitude calculated during the training of a neural network that is used to update the network weights in the right direction and by the right amount. In deep networks or recurrent neural networks, error gradients can accumulate during an update and result in very large gradients. These in turn result in large updates to the network weights, and in turn, an unstable network. At an extreme, the values of weights can become so large as to overflow and result in NaN values. The explosion occurs through exponential growth by repeatedly multiplying gradients through the network layers that have values larger than 1.0

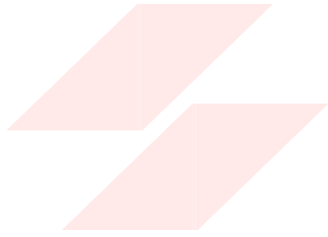
**Vanishing Gradient** The vanishing gradients problem is one example of unstable behavior that you may encounter when training a deep neural network. It describes the situation where a deep multilayer feed-forward network or a recurrent neural network is unable to propagate useful gradient information from the output end of the model back to the layers near the input end of the model.

15. What Is the Difference Between Epoch, Batch, and Iteration in Deep Learning?

**Answer:** The batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters.

The number of epochs is a hyperparameter that defines the number times that the learning algorithm will work through the entire training dataset.

An **iteration** is a term used in **machine learning** and indicates the number of times the algorithm's parameters are updated



**FLIP ROBO**

