

# DEEP LEARNING

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

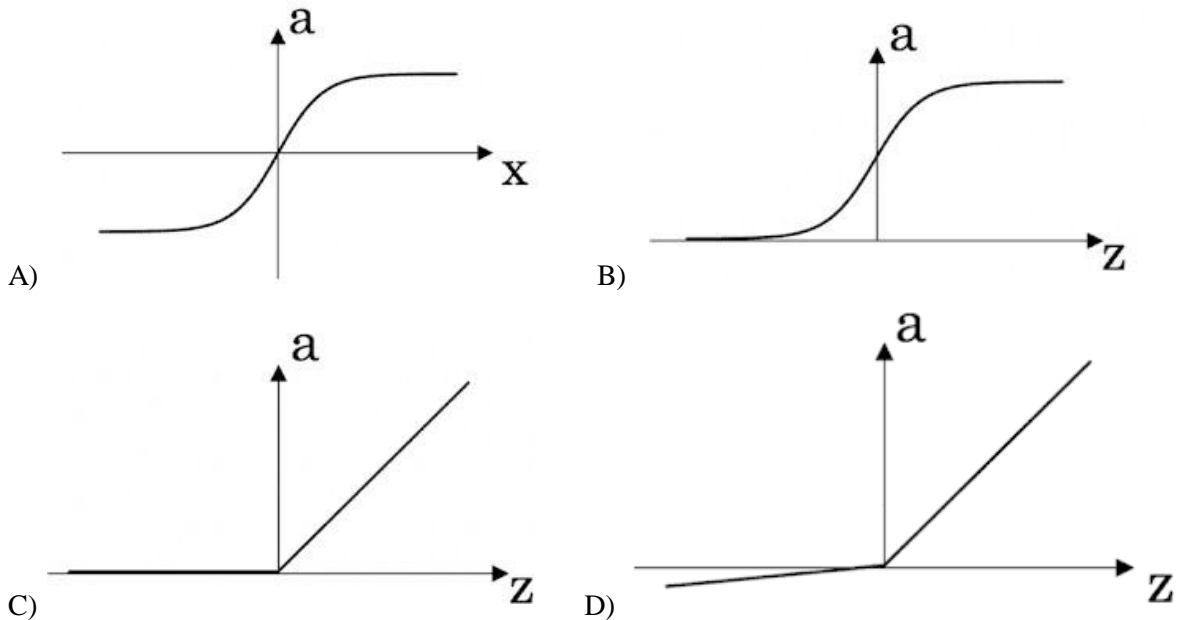
1. Operations in the neural networks can performed\_\_\_\_?
- A) serially  
B) parallelly  
C) serially or parallelly  
D) None of the above

**ANS: C) serially or parallelly**

2. Who proposed the first perceptron model and when?
- A) Rosenblatt, 1958                      B) McCulloch-pitts, 1958  
C) John Hopfield, 1982                D) McCulloch-pitts, 1982

**ANS: A) Rosenblatt, 1958**

3. Which one of these plots represents a ReLU activation function?



**ANS: 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)  $[8 \times 3]$ ,  $[5 \times 3]$

**ANS: A) [3×8], [5×8]**

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

**ANS: C) A unit which doesn't update during training by any of its neighbour**

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

ANS: **B) softmax**

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

ANS: **D) weights**

8. What is asynchronous update in neural networks?
- A) output units are updated parallelly  
B) output units are updated sequentially  
C) either sequentially or parallelly  
D) None of the above

ANS: **B) output units are updated sequentially**

**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

ANS: **A) EarlyStopping, B) Dropout**

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->....

ANS: **B) It is strictly more powerful than a Convolutional Neural Network**

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

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

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

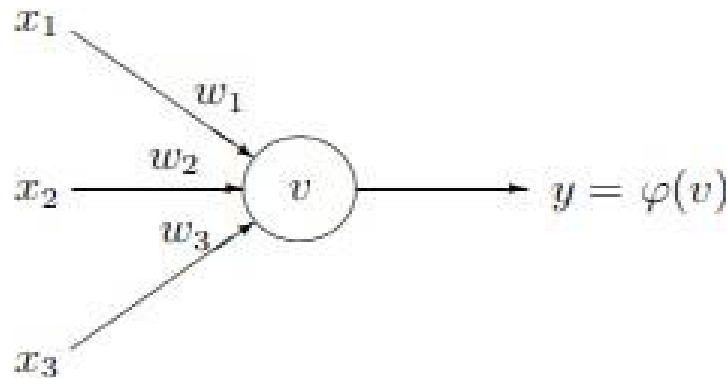
Where  $f(x)$  is the activation function. If you want to build a perceptron which gives an output for linear regression, what will be the activation function you would use?

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

- The learning rate is one of the hyperparameter that plays an important role to decide the complexity of the model being created. The value of hyperparameter decides the complexity of the model.
- The learning rate is a scalar value which is used to decide how accurately the model will converge on a result.
- A very high learning rate will converge the model towards the result rapidly but it will overshoot the minima. During the convergence it'll also miss some important relations and missing this will generate wrong result.
- On the other hand setting a very low learning rate will take a longer time and will capture every details even though they are not required and at the end it will produce wrong results.

- The momentum generated by learning rate can be controlled by using learning rate decay.

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?

Here we have 3 inputs  $x_1, x_2, x_3$  with binary signal so number of input pattern will be as follows;

$x_1$	1	0	0	1	1	0	1	0
$x_2$	0	1	0	1	0	1	1	0
$x_3$	0	0	1	0	1	1	1	0

So, its giving 8 different pattern of inputs

Suppose we have 4 input named  $x_1, x_2, x_3, x_4$  with binary signals 0 and 1 so the input patterns will be as follows;

$x_1$	0	1	1	0	0	0	0	1	1	1	0	0	0	1	1	1
$x_2$	0	1	0	1	0	0	1	0	1	1	0	1	1	1	0	0
$x_3$	0	1	0	0	1	0	1	1	0	1	1	1	0	0	1	1
$x_4$	0	1	0	0	0	1	1	1	1	0	1	0	1	0	0	1

So, 4 inputs with binary signals generates 16 distinct input pattern

So, basing on that 5 inputs with binary signal will give 32 distinct input pattern

We can derive a formula from the above

No of input pattern=  $(2)^{\text{no of inputs}}$

(where 2 represents the binary signals)

No of input patterns at 3 input=  $2^3=08$

No of input patterns at 4 input=  $2^4=16$

No of input patterns at 5 input=  $2^5=32$

#### 14. What Are Vanishing and Exploding Gradients?

- In deep neural network while training with gradient based learning and backpropagation techniques, partial derivatives is obtained by traversing the network from final layer to the initial layer. The derivatives of the inner layer are obtained by matrix multiplication and chain rule.
- So, in a network of n hidden layers, n number of derivatives will be generated and higher value of derivative leads to an exponential increase in the value of the gradient as we propagate and this will explode and it is termed as exploding gradient.
- However, if the values obtained in derivative have lower values the value of the gradient will goes down while propagating and at some point, it will be vanished and it is called as vanishing gradient
- In case of exploding gradient the extremely high value the derivatives lead to instability in the model and made it incapable of effective learning

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

BATCH	EPOCH	ITERATION
The batch size is the number of samples to be processed before the model is updated.	The number of epochs is the number of complete passes through the training dataset.	Iterations is the number of batches needed to complete one epoch.