

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 parallelly

D) None of the above

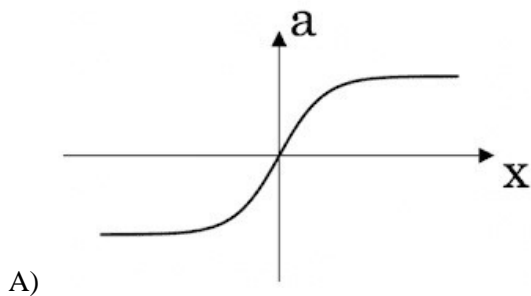
2. Who proposed the first perceptron model and when?

→ A) Rosenblatt, 1958

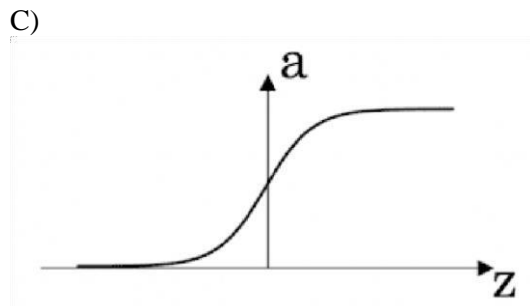
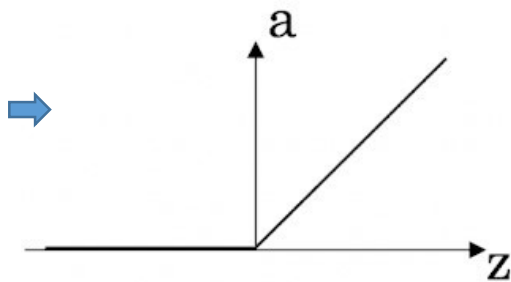
B) McCulloch-pitts, 1958 C) John Hopfield, 1982

D) McCulloch-pitts, 1982

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



B)



D)

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?

B) $[8 \times 3]$, $[5 \times 8]$

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
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
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
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

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
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->....

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?

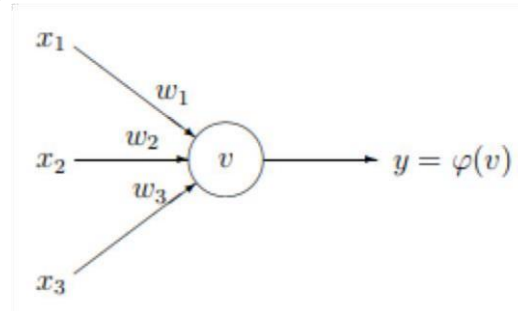
The required activation function will be $f(b + x)$.

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

The learning rate controls how quickly the model is adapted to the problem. Smaller learning rates require more training epochs given the smaller changes made to the weights each update, whereas larger learning rates result in rapid changes and require fewer training epochs.

A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck.

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?

For 3 inputs the number of combinations of 0 and 1 is 8.

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

For 4 inputs, the no. of combinations of 0 & 1 is 16.

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

For 5 inputs, the no. of combinations is 32.

Formula for the number of binary input patterns is 2^n , where n is the number of inputs.

14. What Are Vanishing and Exploding Gradients?

In a network of n hidden layers, n derivatives will be multiplied together. If the derivatives are small then the gradient will decrease exponentially as we propagate through the model until it eventually vanishes, and this is the **vanishing gradient** problem.

If the derivatives are large then the gradient will increase exponentially as we propagate down the model until they eventually explode, and this is what we call the problem of **exploding gradient**.

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

Epoch:

One Epoch is when an entire dataset is passed forward and backward through the neural network only once.

Batch size:

Total number of training examples present in a single batch.

Iteration:

Iterations is the number of batches needed to complete one epoch. (*The number of batches is equal to number of iterations for one epoch.*)