**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 ?
   1. serially B) parallely

C) serially or parallely D) None of the above

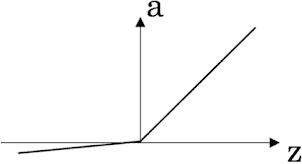
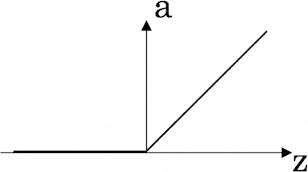
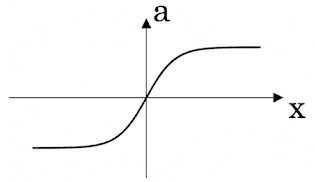
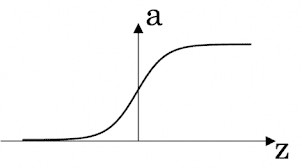
**Answer:** C) serially or parallely

1. Who proposed the first perceptron model and when?
   1. Rosenblatt, 1958 B) McCulloch-pitts, 1958

C) John Hopfield, 1982 D) McCulloch-pitts, 1982

**Answer:** A) Rosenblatt, 1958

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

A)  B) 

C) D)

**Answer:** C)

1. 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) [3×8], [5×8] B) [8×3], [5×8]

C) [5×8], [8×5] D) [8×3], [5×3]

**Answer:** B) [8×3], [5×8]

1. What is a dead unit in a neural network?
   1. A unit which does not respond completely to any of the training patterns
   2. The unit which produces the biggest sum-squared error
   3. A unit which doesn’t update during training by any of its neighbour
   4. None of these

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

1. 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?
   1. sigmoid B) softmax

C) tanh D) ReLU

**Answer:** B) softmax

1. The amount of output of one unit received by another unit depends on what?
   1. output unit B) input unit

C) activation values D) weights

**Answer:** D) weights

1. What is asynchronous update in neural networks?
   1. output units are updated parallely B) output units are updated sequentially

C) either sequentially or parallely D) None of the above

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

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

1. Which of the following techniques can be used to reduce overfitting in a neural network?
   1. EarlyStopping B) Dropout

C) checkpoints D) ReduceLROnPlateau

**Answer:** A) EarlyStopping

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

**Answer:** A) It can be trained as a supervised learning proble, C) It is applicable when the input/output is sequence (e.g., a sequence of words)

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

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

*n*

*y*  *f* (*b*   *wi xi* )

*i* 1

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?

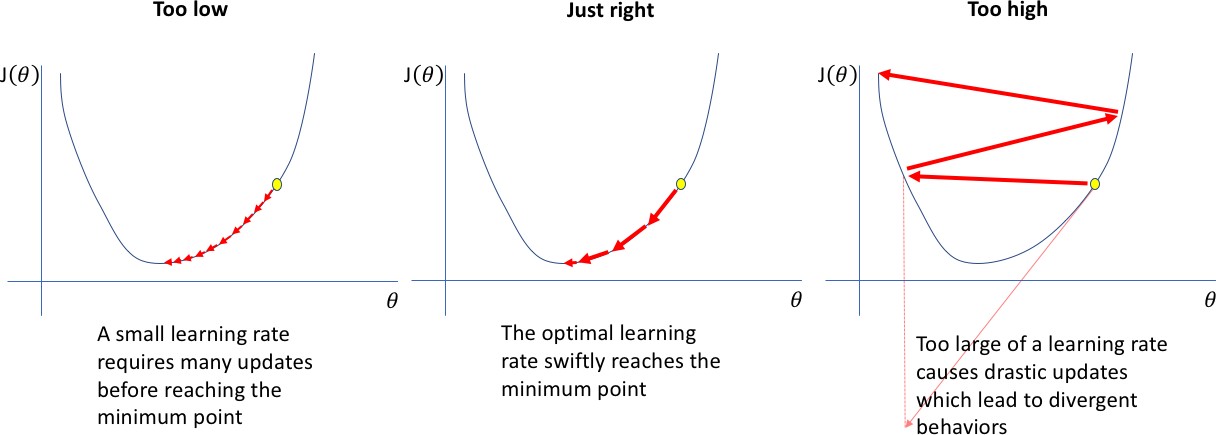
**Answer:** Activation function are used for imparting non-linearity to the output of a simple perceptron. Output of a simple perceptron is sum of bias and weighted sum of all the other variables, which is nothing but the output of a linear regression model. SO, in this case there is no need of an activation function. Therefore, our activation function would be:

f(x)= kx where k is a constant number.



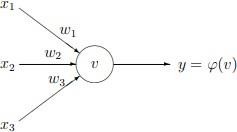
1. What will happen if we use very large or very small learning rates?

**Answer:** If you learning rate is set too low, training will progress very slowly as you are making very tiny updates to the weights in your network. Due to the small updates in the learning rates the loss function decreases at a very slow pace and the algorithm may get stuck at local minima. However, if learning rate is set too high, it can cause undesirable divergent behaviour from minima. Also in such case the learning rate fluctuates and algorithm does not settle at minima. Therefore, adaptive learning rates are used, in initial phases of training the learning rate changes rapidly and as it gets closer to global minima its momentum to change decreases. Refer the following figure for better understanding:



1. Below is a diagram if a single artificial neuron:

WORKSHEET



The node has three inputs x = (x1, x2, x3) 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:** For three inputs the number of combinations of 0 and 1 is 8:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| X2 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| X3 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

and for four inputs the number of combinations is 16:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| X2 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| X3 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| X4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

You may check that for five inputs the number of combinations will be 32. Note that 8 = 23, 16 = 24 & 32 = 25. Thus, the formula for the number of binary input patterns is: 2n, where n in the number of inputs.

1. What Are Vanishing and Exploding Gradients?

**Answer:** Exploding and vanishing gradients’ problem occur in very deep neural networks and because of these problems the training of deep neural networks becomes difficult. In deep multilayer Perceptron networks, exploding gradients can result in an unstable network that at best cannot learn from the training data and at worst results in NaN weight values that can no longer be updated.

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

Answer: **Epoch** - Represents one pass over the entire dataset (everything put into the training model).

**Batch** - Refers to when we cannot pass the entire dataset into the neural network at once, so we divide the dataset into several batches.

**Iteration** - if we have 10,000 images as data and a batch size of 200. then an epoch should complete in 50 iterations (10,000 divided by 200).