**WORKSHEET-2**

**DEEP LEARNING with Answers**

# 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

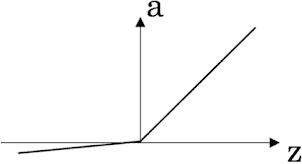
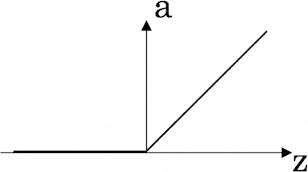
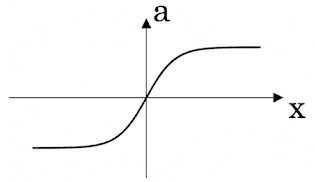
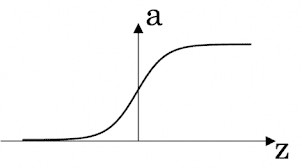
Answer: ( C ) serially or parallely

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)

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

A)  B) 

C) D)

Amswer: ( 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: ( A)

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 )

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: (A )

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)

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)

# 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) (B)

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: (B) (A)

# 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:

Perceptron with a Boolean output:

Inputs: x1…xn

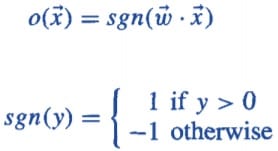
Output: o(x1….xn)



Weights: wi=> contribution of input xi to the Perceptron output;

w0=> bias or threshold

If ∑w.x > 0, output is +1, else -1. The neuron gets triggered only when weighted input reaches a certain threshold value.



An output of +1 specifies that the neuron is triggered. An output of -1 specifies that the neuron did not get triggered.

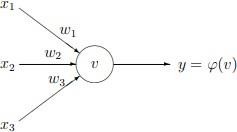
“sgn” stands for sign function with output +1 or -1.

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

Answer: Generally, a **large learning rate** allows the model to learn faster, at the cost of arriving on a sub-optimal final set of weights. A **smaller learning rate** may allow the model to learn a more optimal or even globally optimal set of weights but may take significantly longer to train.

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 = 2 3 , 16 = 2 4 and 32 = 2 5 (for three, four and five inputs). Thus, the formula for the number of binary input patterns is: 2 n , where n in the number of inputs

1. What Are Vanishing and Exploding Gradients?

Answer: The term **vanishing gradient** refers to the fact that in a feedforward network (FFN) the backpropagated error signal typically decreases (or increases) exponentially as a function of the distance from the final layer. — Random Walk Initialization for Training Very Deep Feedforward Networks,

In a network of n hidden layers, n derivatives will be multiplied together. 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** .

**Exploding gradients** are a **problem** where large error **gradients** accumulate and result in very large updates to neural network model weights during training. This has the effect of your model being unstable and unable to learn from your training data.

A

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

Answer: Epochs: An **epoch** is a term used in **machine learning** and indicates the number of passes of the entire training dataset the **machine learning** algorithm has completed. ... If the batch size is the whole training dataset then the number of **epochs** is the number of iterations.

Batch and Epochs:

The **batch** size is a number of samples processed before the model is updated. The number of **epochs** is the number of complete passes through the training dataset. The size of a **batch** must be more than or equal to one and less than or equal to the number of samples **in the** training dataset

Epochs and Iteration:

**Iterations** is the number of batches of data the algorithm has seen (or simply the number of passes the algorithm has done on the dataset). **Epochs** is the number of times a learning algorithm sees the complete dataset.

