DL_NN MOCK TEST

1.You are using the dropouts regularization technique to regularize a neural network. In the given context, how is the mask α for each layer generated?

    A. The mask α is generated independently for each layer during feedforward
    B. The mask α is generated randomly for each layer during feedforward
    C. The mask α for each layer is generated depending on the mask of the previous layer
    D. None of the above

2. The TensorFlow library internally models and stores computations as a directed graph. What do the nodes of the graph represent?

    A. The flow of data
    B. The computations to be performed
    C. Contains some abstract operations which take in some input
    D. Both 2 and 3
    E. Both 1 and 3

3. What will happen if all the layers in a convolutional neural network do convolutions without padding?

    A. It will shrink the output
    B. It will lead to dense layers
    C. It will maintain the size of the output
    D. it will shrink only certain pixels

4. Which of these is a valid way of using pre-trained networks for transfer learning?
1. Freeze the (weights of) initial few layers and training only a few latter layers
2. Retrain the entire network (all the weights) initialising from the learned weights

    A. Only 1
    B. Only 2
    C. Both 1 and 2
    D. None of these

5. Why is the "divide by 255" strategy not used when normalising a CRX dataset?

    A. The CRX images are not natural images
    B. The CRX images are natural images
    C. The range of pixels in CRX images is undetermined
    D. Both 1 and 3
    E. Both 2 and 3

6. Which of the following is NOT a gating mechanism used in LSTMs?
1. the forget gate
2. the update gate
3. the input gate

    A. Only 3
    B. Only 2
    C. Only 1
    D. Both 2 and 3
    E. Both 1 and 2

7. The objective function used to represent the regularization technique is the sum of the Loss function and the regularization term. What does the regularization term in the L2 norm represent?

    A. The sum of all the model parameters
    B. The product of all the model parameters
    C. The sum of squares of all the model parameters
    D. The product of squares of all the model parameters

8. You are given a step function where y =1 if x>0 and y = 0 if x<=0. What will the output of the perceptron be for the given the vectors.
w = [5 4 5 6] x = [1 1 0 0] and bias = - 2

    A. 1
    B. 0
    C. -1
    D. None of the above

9. You are trying to convolve an image using a combination of filter and stride. Which of these combinations cannot be used to convolve an image?

A. A (4, 4) image with a (3, 3) filter using a stride of 2 cannot be convolved
B. A (5, 5) image with a (2, 2) filter and a stride of 2 cannot be convolved
C. Both 1 and 2
D. None of these

10. You have a weight matrix 'Wl' to which you want to apply a mask α. What will be dimension of the mask vector α?

A. (number of neurons in layer 'l-1', 1)
B. (number of neurons in layer 'l', 1)
C. (1, number of neurons in layer 'l-1')
D. (1, number of neurons in layer 'l')

11. You are using Pre- trained Convolutional Neural Network for transfer learning. What information are the initial layers of the network capable of extracting in this scenario?

A. The initial layers discriminate between images
B. The initial layers extract abstract features of an image
C. The initial layers extract detailed representations of an image
D. None of these

12. A long, short-term memory network (LSTM) was designed to solve the vanishing gradient problem in RNNs. Which of these statements is valid with respect to LSTMs?

A. The neurons in the LSTM are all called cells.
B. LSTM has an explicit memory unit which stores information relevant for learning some task..
C. The memory units in LSTMs, retain pieces of information even when the sequences get really long
D. Both 1 and 2
E. Both 1 and 3

13. A RNN consumes a sequence which finishes after T time steps. The output from the last layer of the network captures the entire sequence fed to the network.
What can the output be used for?

    A. It can be used to classify the sentence as correct/incorrect
    B. It can be used to feed it to a regression output
    C. Both 1 and 2
    D. None of these

14. You are trying to compute the output of the ith neuron in the layer I in a feedforward neural network. Which of these is a mandatory step that should be followed before applying the activation function, to get the output of the ith neuron?

    A. Convert the weighted sum to cumulative sum by adding the ith bias term of the bias vector.
    B. Convert the cumulative sum to weighted sum by adding the ith bias term of the bias vector.
    C. Multiply the ith row of the weight matrix with the output of layer I-1 to get the weighted sum of inputs
    D. Both 1 and 3
    E. Both 2 and 3

15. You have two matrices X and Y of size 3x3 as given alongside. You want to convolve the image X using the image Y. What will be the output of performing convolution of X and Y?

X				Y		
1	0	2		4	2	4
4	3	2		2	3	2
0	1	3		2	0	1

    A. 36
    B. 35
    C. 40
    D. 46

16. Consider the image and a 2x2 filter given alongside. What will the convolved image be after using the given filter on it?

3	7	2
10	0	7
0	2	0

-1	0
1	0

A.

-7	7
10	-2

B.

-7	7	0
10	-2	0
1	0	0

C.

-7	7
10	0

D. None of these

---

17. A cell in an LSTM layer receives the following inputs.

1. The output of the previous time step (a vector)

2. The current input (a vector)

3. The previous cell state (a scalar)

In the given context, which of these outputs does the cell produce?

A. The current cell state of the cell (a vector) and the current state output (a vector)

B. The current cell state of the cell (a vector) and the current state output (a scalar)

C. The current cell state of the cell (a scalar) and the current state output (a vector)

D. The current cell state of the cell (a scalar) and the current state output (a scalar)

18. You want to convolve a (6,6) image with a (3,3) filter and a stride 1. Further, you convolve the output with a (3,3) filter. The image is shrinked at the end of two convolutions. What is the size of padding that you can use to avoid shrinkage of the image?

    A. Padding of size 1 pixel should be used on all 4 ends of the matrix.
    B. Padding of size 2 pixel should be used on all 4 ends of the matrix.
    C. Padding of size 3 pixel should be used on all 4 ends of the matrix.
    D. Either 1 or 2
    E. Either 2 or 3

19. Given an input volume of size 1000x1000x64, and we do padding of 1, what will be the output volume?

    A. 1000x1000x65
    B. 1000x1000x66
    C. 1001x1001x64
    D. 1002x1002x64

20. You need to classify 10x10 grey-scale images into one of the 4 classes: cat, dog, bird or none. You decide to use an Artificial Neural Network with 4 hidden ReLU layers having 5, 10, 15 and 21 neurons respectively. Each pair of consecutive layers is densely connected and every neuron in the network carries a bias. The training is done with mini-batches of size 64. The output layer will be a:

    A. Layer having 4 neurons with the ReLU activation function and zero bias in each
    B. Softmax layer with 4 output neurons
    C. Sigmoid layer with 4 output neurons
    D. Layer having 4 neurons with the tanh activation function in each

21. Given an input volume of size 150x150x128 and you apply 10 filters, each of size 5x5 and depth equal to the depth of the input volume. What is the total number of parameters (including bias) in the filters?

    A. 32000
    B. 32010
    C. 28800000
    D. 28800010

22. You are using VGG-16 architecture to build a Convolutional Neural Network in which the first convolutional layer takes a (128, 128, 3) image as the input.
The layer also has 32 filters of size (3, 3, 3).
What are respective weights, biases and trainable parameters for the first convolutional layer in the given context?

    A. Weights : 64 x 3 x 3 x 3 Biases : 64 Trainable parameters: 1792
    B. Weights : 32 x 3 x 3 x 3 Biases : 32 Trainable parameters: 896
    C. Weights : 32 x 3 x 3 x 3 Biases : 32 Trainable parameters: 1792
    D. Weights : 64 x 3 x 3 x 3 Biases : 64 Trainable parameters: 896

23. Which of the following is incorrect about an encoder-decoder RNN architecture?

    A. An encoder-decoder model is a many-to-many type architecture.
    B. The encoder model consumes the input and the decoder model gives the output
    C. The length of the input and output sequences in an encoder-decoder model can be the same
    D. The weight update in the decoder model depends on the gradients from the encoder model

24. In a typical CNN architecture, the height and width of the representation reduces through the layers and the number of channels in the form of feature maps increases.
This makes sense because …
I. Making the representation smaller, reduces redundant information
II. The deeper layers are expected to capture richer features of larger parts of the image
III. This brings down the number of parameters in the network

    A. I and II
    B. I, II and III
    C. II
    D. None of these

25. Which of the following helps to reduce variance in the gradients computed and hence the results in faster convergence?

    A. Batch Normalization
    B. Dropout
    C. L1/L2 Regularisation
    D. None of the above