DNN Answer Key July 2024

BITS WILP M Tech Data Science & Engineering (Birla Institute of Technology and Science, Pilani)

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Work Integrated Learning Programmes Division

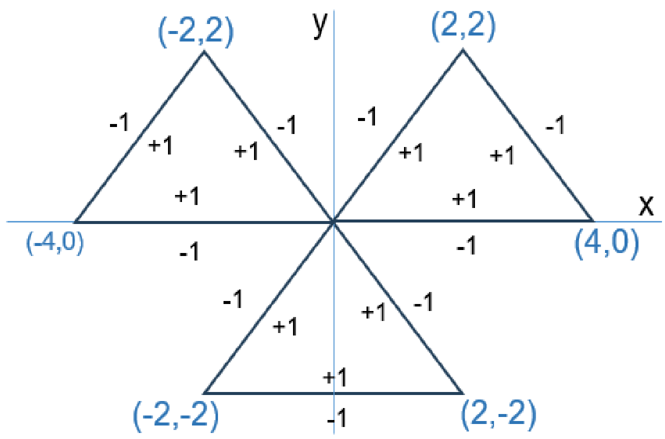
Cluster Programme - M. Tech in AI & ML and DSE

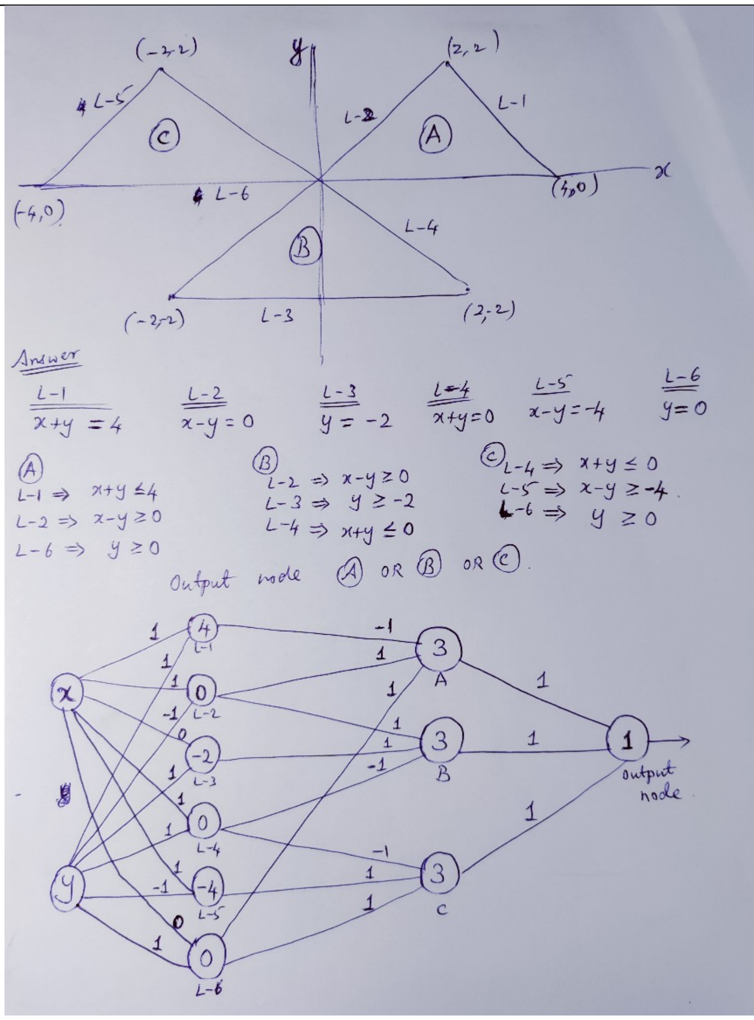
II Semester , 2023 – 24(July,2024)

Mid semester Examination (**MAKEUP)**_ANSWER KEY

Course Title	:	DEEP NEURAL NETWORK / Deep Learning
Nature of Exam.	:	Closed Book
Weightage	:	30 Marks
Duration	:	120 minutes
Date	:	28st July,2024_10 AN

Number of questions:6
Number of Pages: 2

Q. No	Question	Marks
Q.1.	Design a fully connected multilayer perceptron network with minimum number of hidden layers and hidden nodes required to classify the below decision boundary with 100% accuracy. (x, y) are input features and target classes are either +1 or -1 as shown in the figure.Step activation functions are used at all nodes, i.e., output=+1 if total weighted input >= threshold (mention inside node) at a node, else output = -1. 	**6M**
SOL	6 neurons in first layer, 3 neurons in 2nd layer, 1 neuron in last layer	



Q.2. (a) Given an input matrix of size 8×8 and a kernel size of 3×3 with a stride of 1 and no padding, determine the output shape after applying the convolution operation. [2 M]

(b) If there are 5 such filters applied, what will be the output shape? [1 M]

(c) Apply a 4×4 max-pooling filter with a stride of 2 to the output from part (b). What will be the final output shape? [2 M]

5M

SOL

(a) Given:

- Input size N=8N = 8N=8,
- Filter size F=3F = 3F=3,
- Padding P=0P = 0P=0,
- Stride S=1S = 1S=1.

Plugging these values into the formula:
Output dimension=(8−3+2⋅0)/1+1=5/1+1=6
So, the output shape after the convolution is 6×6
(b) Since there are 5 such filters applied, the number of output channels

	will be 5. Thus, the output shape will be: 6×6×5 (c) Output dimension=(6−4)/2+1=2/2+1=2 Since we have 5 channels, the final output shape will be: 2×2×5	
Q.3.	Mr Ram has a dataset **data_1.csv** with 1 million labelled training examples for classification, and dataset **data_2.csv** with 100 labelled training examples. Mr Rakesh trains a model from scratch on **data_2.csv**. Mr Ram decided to train on **data_1.csv**, and then apply transfer learning to train on **data_2.csv**. Differentiate between these approaches and the advantages / problems both of them faces and how to solve them, if possible? State one problem Mr Rakesh is likely to find with his approach. How does Mr Ram approach address this problem?	**4M**
Q.3.	Mr Rakesh is likely to see overfitting. Model is not going to generalise well to unseen data. By using transfer learning and freezing the weights in the earlier layers, Mr Ram reduce the number of learnable parameters, while using the weights which have been pretrained on a much larger dataset.	
Q.4.	Given an error function $E(w_1,w_2) = 3w_1^2 + 4w_2^2 + 2w_1w_2$,different variants of gradient descent can be used to minimize the error with respect to w_1 and w_2. Assume initial weights are $(w_1, w_2) = (0.5,0.5)$ at time t−1. (a) Calculate the gradients ∂E/∂w1 and ∂E/∂w2. [3 M] (b) Using the standard gradient descent with a learning rate η=0.1, compute the weight updates and the new weights at time t. [2 M]	**5M**
	(a) ∂E/∂w1=6w1+2w2=4 ∂E/∂w2=8w2+2w1=5 (b) w1=0.5-0.1*4=0.1 w2=0.5-0.1*5=0	
Q.5	a) Discuss any two benefits of using convolution layers instead of fully connected layers for image classification [2 M] b) Suppose that you are training a deep learning algorithm on a given	**5M**

	data set. You observed that accuracy of the algorithm is decreasing after few epochs. Then what is your interpretation of this and how to address this situation, if possible [3 M]	
	1. Enhanced Feature Extraction [1 marks] Convolutional layers excel in extracting features from images due to their ability to learn localized patterns, such as edges and textures. This is achieved through the use of small filters or kernels that scan over the input image, allowing the network to capture spatial hierarchies in a hierarchical manner as the layers progress. As a result, convolutional layers can gradually learn more complex features, leading to improved performance in tasks such as image classification and object recognition. **2. Computational Efficiency [1 marks]** When compared to fully connected layers, convolutional layers are significantly more efficient in terms of computational requirements. This efficiency arises from their use of sparse connections and weight sharing, which drastically reduces the number of parameters that need to be trained. Fully connected layers, on the other hand, involve dense connections, resulting in many more parameters, leading to higher computational costs and a greater risk of overfitting. Thus, convolutional layers facilitate the processing of high-dimensional image data effectively. **Solution b)** **Interpretation of Decreasing Accuracy [1.5 marks]** When observing a decrease in accuracy after a few epochs while training a deep learning algorithm, this often indicates that the model is beginning to overfit the training data. Overfitting occurs when the model learns the noise and specific details of the training set instead of generalizing well to unseen data. This results in good performance on the training set but poor performance on the validation or test set, signified by a rising error in those datasets while the training accuracy continues to improve.	

	Approaches to Address Overfitting [1.5 marks] 1. **Early Stopping** 2. **Regularization Techniques** 3. **Using Dropout Layers** 4. **Data Augmentation** 5. **Adjusting Model Complexity** 6. **Hyperparameter Tuning**	
Q.6.	a) List the hyper parameters used for training a neural network. [2 M] b) We know that ANN help us in image classification. If so, why CNN is preferred over ANN for image classification [3 M]	**5M**
	a) 1. Number of layers 2. Number of nodes in a layer 3. Weights and biases 4. Activation functions b) Number of trainable parameters are more and hence the computational complexity increases. ANN ignores spatial information etc	