Master of Computer Applications
Deep Learning
March 2023

Time: One hour **Max. Marks: 35**

Note: ML denotes machine learning throughout the question paper.

1.	For each of the following questions, answer in 3-4 lines	7x3=21

a. You design a fully connected neural network architecture where all activations are sigmoids. You initialize the weights with large positive numbers. Is this a good idea? Justify your answer.

b. You are given a dataset of 10 × 10 grayscale images. Your goal is to build a 5-class classifier. Which of the following approaches is the better option (if any):
- The input is flattened into a 100-dimensional vector, followed by a fully-connected layer with 5 neurons
- The input is directly given to a convolutional layer with five 10 × 10 filters

Justify your answer.Justify

c. You'd like to train a fully-connected neural network with 5 hidden layers, each with 10 hidden units. The input is 20-dimensional and the output is a scalar. What is the total number of trainable parameters in your network?

d. Consider the following sigmoid and tanh activation functions:

$$\sigma(z) = \frac{1}{1+e^{-z}} \text{ and } tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}$$

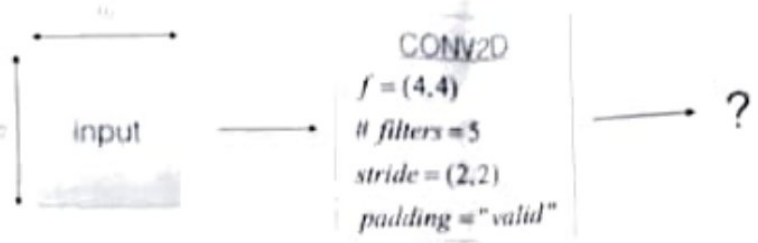
Calculate $\partial\sigma(z)/\partial z$ and $\partial\tanh(z)/\partial z$.

e. Assume that before training your neural network, the setting is:
- The data is zero centered.
- All weights are initialized independently with mean 0 and variance 0.001.
- The biases are all initialized to 0.
- Learning rate is small and cannot be tuned.

Using the result from (c), justify which activation function between tanh and sigmoid is likely to lead to a higher gradient during the first update

f. Give the diagram of an LSTM unit, showing the relevant equations.

g. How can you address the gender bias in word embeddings?

2.	State one advantage and one disadvantage of pooling when CNN is applied to images. Consider the figure below  Figure 1: Input of shape $(n_H, n_W, n_C) = (10, 10, 1)$; There are five 4×4 convolutional filters with 'valid' padding and a stride of (2, 2) What is the output shape after performing the convolution step in Figure 1? Write your answer in the following format: (n_H, n_W, n_c).	(3+4=7)
3.	Derive a gradient descent training rule for a two layer network with a single output unit employing sigmoid activation function. The network is supplied with m training instances. The intermediate layer employs ReLu activation function. The network uses the cost function for logistic regression given by: `E(t,y) = -(ylog(t) + (1-y) log(1-t))` where t is the target output and y is the neural network output.	(7)