JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATIONS- 2026

M.Tech.-II Semester (CSE)

COURSE CODE (CREDITS): 22M11CI212 (03) MAX MARKS: 35

COURSE NAME: DEEP LEARNING TECHNIQUES

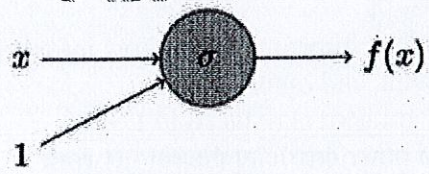
COURSE INSTRUCTOR: RBT MAX. TIME: 2 Hours

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

(c) Use of calculator is allowed

Q.No	Question	CO	Marks
Q1	a) Consider a recurrent network in which the hidden states have a dimensionality of 2. Every entry of the 2×2 matrix W_{hh} of transformations between hidden states is 3.5. Furthermore, sigmoid activation is used between hidden states of different temporal layers. Would such a network be more prone to the vanishing or the exploding gradient problem? b) Recommend possible methods for pre-training the input and output layers in the machine translation approach with sequence-to-sequence learning. **OR** a) Draw and explain LSTM architecture. Show how information flows through forget, input, and output gates. b) Explain how vanishing gradient is reduced in LSTM.	3	2.5 + 2.5 = 5
Q2	How is RNN different from other artificial neural network architectures?	3	5
Q3	Consider two neural networks used for regression modeling with identical structure of an input layer and 10 hidden layers containing 100 units each. In both cases, the output node is a single unit with linear activation. The only difference is that one of them uses linear activations in the hidden layers and the other uses sigmoid activations. Which model will have higher variance in prediction?	4	5
Q4	a) Suppose that you have a model that provides around 80% accuracy on the training as well as on the out-of-sample test data. Would you recommend increasing the amount of data or adjusting the model to improve accuracy? b) How does backpropagation work in CNN?	4	5
Q5	Consider a network with a single input layer, two hidden layers, and a	2	5

	single output predicting a binary label. All hidden layers use the sigmoid activation function and no regularization is used. The input layer contains d units, and each hidden layer contains p units. Suppose that you add an additional hidden layer between the two current hidden layers, and this additional hidden layer contains q linear units. (a) Even though the number of parameters have increased by adding the hidden layer, discuss why the capacity of this model will decrease when $q > p$? (b) Does the capacity of the model increase when q > p?		
Q6	a) What is the effect of increasing the regularization parameter on the training and testing accuracy? At what point do training and testing accuracy become similar? b) What is the Attention Mechanism? Why was it introduced? Explain the working of attention mechanism with diagram. **OR** a) How do attention mechanisms improve LSTM? b) Why are Transformers replacing LSTMs?	2	2.5 + 2.5 = 5
Q7	a) Does the classification accuracy on the training data generally improve with increasing training data size? How about the point-wise average of the loss on training instances? At what point do training and testing accuracy become similar? Explain your answer. b) Consider the following computation,  $f(x) = tanh(w \cdot x + b)$ The value L is given by, $$L = \frac{1}{2}(y - f(x))^4$$ Here, x and y are constants and w and b are parameters that can be modified. In other words, L is a function of w and b. Derive the partial derivatives, ∂L/∂w and ∂L/ ∂b	1	2.5 + 2.5 = 5