REG.NO.: 043

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

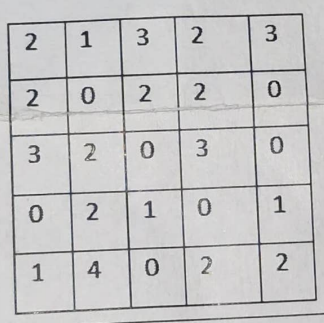
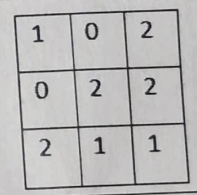
SLOT: A1+TA1

Programme Name & Branch	: B. Tech
Course Code and Course Name	: BCSE332L and Deep Learning
Faculty Name(s)	: Dr. Vijayanand C
Class Number(s)	: VL2025260502243
Date of Examination	: 15/03/2026
Exam Duration	: 90 minutes

Maximum Marks: 50

General instruction(s): Answer All Questions

- Course Outcomes:
  1. Design and develop custom Deep-nets for human intuitive applications.
  2. Design of test procedures to assess the efficiency of the developed model5

Q. No	Question	Marks	CO	BL
1.	The pixel values of the input image are A. Apply padding of size 1 and perform a convolution process with a stride of 2 using the kernel B. Reduce the dimension of the resultant extracted feature using Max pooling of size 2x2. The values of A and B are A =  B = 	10	3	3
2.	You are a Lead AI Engineer at a startup building a real-time defect detection system for a high-speed manufacturing line. The factory takes high-resolution images of products moving on a conveyor belt. You have the following constraints and requirements: 1. Latency: The system must process images in near real-time (very low inference time). 2. Accuracy: You need to detect both large surface scratches and very tiny microscopic cracks. 3. Hardware: The deployment will be on an edge device with limited memory, but you have a powerful server cluster for the initial training. Which architecture (AlexNet, ResNet, or InceptionNet) would you choose for this task, and why? Justify your answer in detail.	10	4	4

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

SLOT: A1+TA

3.	For the RNN, perform the forward pass and backward propagation through time with the data given below: Parameter — Value Input sequence length (T) — 3 time steps Input size (x) — Scalar (1D) Hidden state size — Scalar (1D) Input values — $x_1 = 0.9, x_2 = 0.6, x_3 = -1.1$ Initial hidden state — $h_0 = 0.0$ Input weight — $W_x = 0.6$ Hidden weight — $W_h = 0.7$ Output weight — $W_y = 1.0$ Bias (hidden) — $b_h = 0.1$ Bias (output) — $b_y = 0.0$ Activation function — tanh True output labels — $y_1 = 0.5, y_2 = 0.8, y_3 = -0.5$ Loss function — Mean Squared Error (MSE) Learning rate (η) — 0.1 Show the values of the hidden states, output, loss per time step and gradients.	10	3	4
4.	You are building a Named Entity Recognition (NER) system for a news aggregator. The system must identify "Companies" and "People." Consider the following two sentences that the AI must process: 1. *"**Apple** is a delicious fruit that grows in temperate climates."* 2. *"**Apple** released its latest quarterly earnings report today."* In both sentences, the word **"Apple"** is the very first token. Analyze why a Standard Unidirectional RNN would likely struggle to correctly classify "Apple" in the first sentence versus the second sentence as it processes the text in real-time. Explain how a BERT architecture solves this specific ambiguity.	10	4	3
5.	You are training an AI to act as a Document Summarizer for legal archives. The model is currently reading a 500-word paragraph about a specific person. The paragraph begins: *"Mr. Henderson, who served as the primary legal counsel for the firm for over thirty years and was known for his meticulous attention to detail during the merger of 1994, finally decided to..."*. At the very end of the paragraph (token #500), the model must predict the correct pronoun to complete the sentence: *"...finally decided to submit _ resignation."* Analyze why a Standard RNN would likely fail to predict "his" correctly in this scenario, even if it had seen the name "Mr. Henderson" at the start. Explain the specific mechanism within an LSTM cell that allows it to retain the information about "Mr. Henderson" over hundreds of intervening words. Compare the gradient flow in a Standard RNN versus an LSTM during the training process for this specific long-range dependency.	10	4	3