

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2026

B.Tech-V Semester (CSE-AIML)

COURSE CODE (CREDITS): 25B11CI511(3)

MAX. MARKS: 15

COURSE NAME: DEEP LEARNING

COURSE INSTRUCTOR: PROF. VIVEK KUMAR SEHGAL

MAX. TIME: 1 Hour

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) What is representation learning, and why is it useful? Discuss the autoencoder as an example of a representation learning algorithm, explaining the roles of the encoder and decoder. (b) Explain how deep learning solves the central problem of representation learning. Describe, with reference to Figure shown below, how a deep learning system might build up the concept of an image of a person from simpler concepts like edges, contours, and object parts. 	CO-1	5
Q2	Suppose you are training a miniature neural network with only two parameters, $\theta = [\theta_1, \theta_2]^T$. The local behavior of your network's loss function $L(\theta_1, \theta_2)$ near a specific point $\theta^* = [0, 0]^T$ is modeled by the following quadratic function: $L(\theta_1, \theta_2) = 3\theta_1^2 - 2\theta_1\theta_2 + 3\theta_2^2$ Answer the following three analytical questions to map the geometry of this loss landscape.	CO-1,2	5

	1. Analytical Derivation: Derive the gradient vector $\nabla L(\theta)$ and find the full 2×2 Hessian matrix H for this loss function. Is the Hessian constant across the entire landscape? Explain why or why not. 2. Eigenvalue Analysis: Calculate the eigenvalues (λ_1, λ_2) of your calculated Hessian matrix. 3. Geometric Interpretation: Based on the signs of the eigenvalues you found, classify the critical point $\theta^* = [0,0]^T$ (e.g., local minimum, local maximum, or saddle point). Explain how an optimizer like Gradient Descent would behave if it were traveling along the path of the sharpest curvature versus the gentlest curvature.		
Q3	A neural network has: $w_0 = 2$ and the gradient is: $g = 0.5$ Learning rate: $\eta = 0.1$ Calculate the first weight update using: 1. SGD 2. Momentum, $\beta = 0.9$ 3. AdaGrad 4. RMSProp, $\beta = 0.9$ 5. Adam, $\beta_1 = 0.9, \beta_2 = 0.999$ Assume all optimizers start with zero accumulated statistics and $\epsilon = 10^{-8}$ extremely small i.e. almost negligible	CO-2	5