180

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATIONS- 2026

M.Tech-II Semester (CSE/IT)

COURSE CODE (CREDITS): 22M11CI212 (03) MAX MARKS: 25

COURSE NAME: DEEP LEARNING TECHNIQUES

COURSE INSTRUCTOR: RBT MAX. TIME: 1 Hour 30 Min

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) ReLU is generally preferred over tanh in neural networks. (True/ False) b) Show the following properties of the sigmoid and tanh activation functions (denoted by $\Phi(\cdot)$ in each case): i. hard tanh activation: $\Phi(-v)=-\Phi(v)$ ii. Sigmoid activation: $\Phi(-v)=1-\Phi(v)$ c) Show that the derivative of the tanh activation function is at most 1, irrespective of the value of its argument. At what value of its argument does the tanh activation take on its maximum value?	1	1 + 2 + 2
Q2	a) A neural network with no non-linear activation functions (i.e., only linear transformations) is equivalent to a single-layer neural network. (True / False). b) Consider a neural network with three layers including an input layer. The first (input) layer has four inputs x_1 , x_2 , x_3 , and x_4 . The second layer has six hidden units corresponding to all pairwise multiplications. The output node o simply adds the values in the six hidden units. Let L be the loss at the output node. Suppose that you know that $\partial L/\partial o = 2$, and x1 = 1, x2 = 2, x3 = 3, and x4 = 4. Compute $\partial L/\partial xi$ for each i.	3	1 + 4
Q3	a) Explain why autoencoders are used for dimensionality reduction. b) The linear autoencoder is applied to each d-dimensional row of the n × d data set D to create a k-dimensional representation. The encoder weights contain the k × d weight matrix W and the decoder weights contain the d × k weight matrix V . Therefore, the reconstructed representation is DW^TV^T, and the aggregate loss value $\|DW^TV^T - D\|^2$ is minimized over the entire training data set. For a fixed value of V, show that the optimal matrix W must satisfy $D^TD(W^TV^TV - V) = 0$.	3	2+ 3

Q4	a) Suppose that the augmented matrix for a linear system has been reduced by row operations to the given row echelon form. Solve the system. $$\begin{bmatrix} 1 & -5 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ b) Find a singular value decomposition of A. $$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$ **OR** Show that if multinomial logistic regression is applied to a data set with k= 2 classes, the resulting updates are equivalent to logistic regression updates.	3	1 + 4
Q5	a) Write advantages of momentum based gradient descent. Explain mathematically how momentum based gradient descent is better than normal gradient descent? b) Let D be an n × d data matrix with non-negative entries. Show how you can approximately factorize $D \approx U V^T$ into two non-negative matrices U and V, respectively, by using an autoencoder architecture with d inputs and outputs. **OR** a) Compare RMSProp, AdaDelta, and Adam gradient descent strategies. b) Consider the softmax activation function in the output layer, in which real-valued out puts $v_1 \ldots v_k$ are converted into probabilities as follows: $$o_i = \frac{\exp(v_i)}{\sum_{j=1}^{k} \exp(v_j)} \quad \forall i \in \{1, \ldots, k\}$$ Show that the value of $\partial o_i / \partial v_j$ is $o_i (1 - o_i)$ when i = j. In the case that i ≠ j,show that this value is $- o_i o_j$.	3	2.5 + 2.5