

**DEEP LEARNING
(AML3202)**

Time Allotted : 2½ hrs

Full Marks : 60

Figures out of the right margin indicate full marks.

*Candidates are required to answer Group A and
any 4 (four) from Group B to E, taking one from each group.*

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve:

12 × 1 = 12

Choose the correct alternative for the following

- (i) The amount of output of one node received by another node depends on what?
 - (a) Number of training samples
 - (b) Number of hidden nodes
 - (c) Weight associated between the given nodes
 - (d) Output of activation function
- (ii) What is the range of output values for tanh activation function?
 - (a) (-1, 1)
 - (b) [-1, 1]
 - (c) {0, 1}
 - (d) {-1, 1}
- (iii) What is the role of gradient descent in deep neural network?
 - (a) To minimize the loss function
 - (b) To find the optimal weights for the neural network
 - (c) To update the network's parameters
 - (d) All of the above
- (iv) What is the vanishing gradient problem?
 - (a) When the gradient value in a deep neural network become extremely small
 - (b) When the gradient value in a deep neural network become extremely large
 - (c) When the gradient value in a shallow neural network become extremely small
 - (d) When the gradient value in a shallow neural network become extremely large
- (v) What is the purpose of a recurrent neural network?
 - (a) deal with sequential data
 - (b) classify images
 - (c) perform dimensionality reduction
 - (d) generate synthetic data
- (vi) What is the purpose of dropout in a deep Boltzmann machine?
 - (a) To deal with imbalanced training data
 - (b) To reduce training complexity
 - (c) To increase the capacity of the network
 - (d) To reduce overfitting

- (vii) Which deep model architecture can be used for anomaly detection and data reconstruction?
 (a) Generative adversarial network (b) Hopfield network
 (c) Deep Boltzmann machine (d) Autoencoder
- (viii) Which deep model architecture is used for reinforcement learning tasks?
 (a) Convolutional neural networks (b) Recurrent neural networks
 (c) Autoencoders (d) Deep Q-Networks
- (ix) What is an agent in reinforcement learning?
 (a) Agent is the situation in which rewards are being exchanged
 (b) Agent is the simple value in reinforcement learning
 (c) An agent is an entity that explores the environment
 (d) None of the above
- (x) In object detection, what does the term *bounding box* refer to?
 (a) The boundary between training and testing data
 (b) A rectangular box that defines the location of an object
 (c) The cropping area for the background
 (d) The area where the loss is minimized

Fill in the blanks with the correct word

- (xi) In case of linear classifier, the decision boundaries are defined by __ dimensional hyperplane within the d-dimensional input space.
- (xii) The parameters of reinforcement network are learned through _____.
- (xiii) _____ CNN architectures were built on ImageNet dataset.
- (xiv) Tokenization converts text into _____ that can be input into the LLM.
- (xv) In a Bidirectional RNN, the hidden states are computed using both _____ and _____ context.

Group - B

2. (a) Discuss the effect of learning rate on the learning of single layer perceptron using back-propagation algorithm. [[CO1](Understand/LOCQ)]
- (b) What would happen to a neural network's performance if bias terms were removed? Explain with reasoning. [[CO1](Understand/LOCQ)]
- (c) Explain the purpose of the momentum term that is often included in the back-propagation learning algorithm. [[CO2](Analyse/IOCQ)]
- (d) A logistic regression model is used to classify whether a student passes an exam ($Y = 1$) or fails ($Y = 0$) based on the number of hours studied (X). The fitted model is as follows:

$$\log \left(\frac{p}{1-p} \right) = -4 + 1.5X$$

where, $p = P(Y = 1|X)$.

A student is classified as Pass ($Y = 1$) if $p \geq 0.5$; otherwise the student is classified as Fail.

Now, answer the following questions:

- (i) Find the probability that a student passes if they studied for 5 hours.
- (ii) Determine the minimum number of study hours required for a student to be classified as Pass.

[[CO3](Apply/IOCQ)]

$$2 + 2 + 2 + (3 + 3) = 12$$

3. (a) Explain the role of the chain rule in back-propagation. [[CO2](Understand/LOCQ)]
- (b) Why does the gradient for earlier layers depend on gradients of later layers? [[CO3](Analyse/IOCQ)]
- (c) Consider the following specifications of a MLP with one input layer, one hidden layer and one output layer:

Inputs: $x_1=1, x_2=0$, Hidden layer (2 neurons, sigmoid), Output layer (1 neuron, sigmoid)

Inputs: $x_1=1, x_2=0$, Hidden layer (2 neurons, sigmoid), Output layer (1 neuron, sigmoid)

Weights: Hidden layer: $w_{11}=0.1, w_{12}=0.3, w_{21}=0.2, w_{22}=0.4$

Output layer: $v_1=0.7, v_2=0.5$

Biases are **zero**, Target $y=1$, Loss: Mean Squared Error.

Now, answer the following questions:

- (i) Perform forward propagation.
- (ii) Compute output error.
- (iii) Compute gradients for output layer weights.
- (iv) Compute gradients for hidden layer weights.
- (v) Update all weights using $\eta=0.1$.

[[CO3](Apply/IOCQ)]

$$2 + 2 + (2 + 1.5 + 1 + 1 + 2.5) = 12$$

Group - C

4. (a) Given the input matrix and the kernel, perform convolution with stride being 1.

1	0	1	1	0
0	0	0	1	1
1	0	0	0	1
0	1	1	1	0
1	1	0	1	0
Input Matrix				

1	0	0
0	0	1
1	1	0
Kernel Matrix		

[[CO3](Apply/IOCQ)]

- (b) Apply max-pooling and sum-pooling on the results from the above convolutions with stride = 1.

[[CO3](Apply/IOCQ)]

$$4 + (4 + 4) = 12$$

5. (a) How can you differentiate between the Attention and Self – Attention technique with respect to the Transformer model. [[CO4](Understand/LOCQ)]
- (b) Explain in detail the role of Self – Attention used in the Transformer model. What benefits can you get by using the Self – Attention technique? [[CO4](Understand/LOCQ)]
- (c) What is the use of the masked self - attention in the decoder segment of the Transformer model? Explain with suitable example. [[CO4](Understand/LOCQ)]

[[CO4](Understand/LOCQ)]

$$2 + (4 + 2) + 4 = 12$$

Group - D

6. (a) Explain the concept of contrastive divergence. [[CO3](Understand/LOCQ)]
 (b) Write down the mathematical equation for energy $E(\mathbf{v}, \mathbf{h})$ of a given state vector $(\mathbf{v}, \mathbf{h})$ in the Boltzmann machine. [[CO3](Apply/HOCQ)]
 (c) What are the major limitations of the learning algorithm of Boltzmann machines? [[CO3](Remember/LOCQ)]
4 + 4 + 4 = 12
7. (a) Write down the energy function for deep Boltzmann machine. [[CO3](Understand/LOCQ)]
 (b) What is the difference between visible and hidden units in deep Boltzmann machines? [[CO3](Remember/HOCQ)]
 (c) Describe the bottom-up forward pass and top-down backward pass in the learning algorithm of deep Boltzmann machines. [[CO3](Analyse/IOCQ)]
4 + 2 + 6 = 12

Group - E

8. (a) Explain Q-Learning. [[CO3](Understand/LOCQ)]
 (b) Explain the differences between the positive and negative reinforcement learning. [[CO5](Analyse/IOCQ)]
 (c) State and explain the steps involved in the reinforcement learning algorithm. [[CO3](Understand/LOCQ)]
4 + 4 + 4 = 12
9. (a) Explain the "Temporal Credit Assignment Problem" in the context of reinforcement learning. [[CO3](Understand/LOCQ)]
 (b) Discuss the trade-off between exploration and exploitation. [[CO3](Understand/LOCQ)]
 (c) Compare Policy Iteration and Value Iteration. Under what conditions does Value Iteration converge faster, and why is Policy Iteration often considered more "stable" in early iterations? [[CO5](Understand/IOCQ)]
3 + 3 + (2 + 2 + 2) = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	48.95	43.75	7.3