

Reg No.: \_\_\_\_\_

Name: \_\_\_\_\_

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**  
B.Tech Degree S7 (S,FE) Examinations May 2026 (2019 Scheme)

**Course Code: AIT401**

**Course Name: FOUNDATIONS OF DEEP LEARNING**

**Max. Marks: 100**

**Duration: 3 Hours**



**PART A**

*Answer all questions, each carries 3 marks.*

Marks

- |    |                                                                                                            |     |
|----|------------------------------------------------------------------------------------------------------------|-----|
| 1  | How does sigmoid and tanh activation function contribute to vanishing gradient problem?                    | (3) |
| 2  | What is the difference between parameters and hyperparameter in a neural network?                          | (3) |
| 3  | Can we apply data augmentation blindly? Justify.                                                           | (3) |
| 4  | How does dropout act as an ensemble model?                                                                 | (3) |
| 5  | Spot the variation between Max Pooling and Average Pooling with example                                    | (3) |
| 6  | What are the different datatypes used in CNN?                                                              | (3) |
| 7  | Discuss the primary objectives of RNN?                                                                     | (3) |
| 8  | What's so special about BERT compared to its predecessors?                                                 | (3) |
| 9  | Describe the use of an over-complete encoder in neural networks and provide an example of its application. | (3) |
| 10 | Why is GANs called Generative 'Adversarial' Networks?                                                      | (3) |

**PART B**

*Answer any one full question from each module, each carries 14 marks.*

**Module I**

- |    |                                                                                                      |      |
|----|------------------------------------------------------------------------------------------------------|------|
| 11 | a) What are the different deep learning techniques used and how can they be classified respectively? | (10) |
|    | b) What is overfitting and how can it be addressed by adding a penalty? Where is the penalty added?  | (4)  |

**OR**

- |    |                                                                         |      |
|----|-------------------------------------------------------------------------|------|
| 12 | a) What are the different activation functions used in neural networks? | (10) |
|    | b) What is the main purpose of Validation Sets?                         | (4)  |

**Module II**

- |    |                                                                                    |     |
|----|------------------------------------------------------------------------------------|-----|
| 13 | a) Explain Gradient descent. Differentiate Gradient descent with Momentum and NAG. | (9) |
|----|------------------------------------------------------------------------------------|-----|

- b) How does the vanishing gradient problem impact the learning process in a neural network? Illustrate this issue using a basic network diagram. (5)

**OR**

- 14 a) What is a cost function? Point out the difference between saddle points, local minimum and maximum, global minimum and maximum with help of a graph. (10)
- b) What is early stopping? (4)

**Module III**

- 15 a) Explain what pre-trained CNN models are with the help of VGG-19. (10)
- b) What is Padding used for and what are the different kinds of Padding? (4)

**OR**

- 16 a) What are the different variants of convolution functions? (8)
- b) Miss C is working with a lemon-orange image dataset, where each image has dimensions  $32 \times 32 \times 3$ . She applies the following operations in sequence: (6)
- i. A convolution operation with a  $5 \times 5$  kernel, 8 filters, and stride 1.
  - ii. A  $2 \times 2$  non-overlapping max pooling operation
  - iii. Another convolution operation with a  $5 \times 5$  kernel, 8 filters, and stride 1.

What is the dimension of the feature map after these operations?

**Module IV**

- 17 a) Explain how LSTM can be used to predict whether a text is spam or not (10)
- b) How does training work in RNN? (4)

**OR**

- 18 a) How does Image captioning work with respect to encoder-decoder RNN (9)  
(Image Captioning : Giving captions to the given image i/p: image , o/p: caption)
- b) Explain the Recursive Neural Network and Deep RNN (5)

**Module V**

- 19 a) What is Autoencoder? Differentiate between Denoising AE and Stochastic AE (8)
- b) Explain the structure and idea behind Boltzmann Machine (6)

**OR**

- 20 a) Which type of neural network is composed of multiple layers of Restricted Boltzmann Machines and is used for unsupervised pre-training of deep learning models? Explain the two ways to train this network. (7)
- b) What does Contrastive Divergence mean in RBM ? (7)

\*\*\*\*