DEEP LEARNING
(CSEN 4142)

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) In deep learning, "feature representation" refers to:
(a) The raw data collected from sensors
(b) The way input data is transformed into meaningful patterns by the network
(c) The visualization of model outputs
(d) The labels of the dataset

(iii) What is the purpose behind using backpropagation in deep convolutional neural network?
(a) To propagate error from the input layer to the output layer
(b) To propagate input activation from the input layer to the output layer
(c) To propagate error from the output layer to the input layer
(d) To propagate output response from the output layer to the input layer

(iv) What is the purpose of a convolutional layer in a convolutional neural network?
(a) To reduce the dimensionality of the input
(b) To extract spatial local features from the input
(c) To classify the input data
(d) To apply non-linear transformations to the input

(v) The "attention mechanism" in Transformers helps the model to:
(a) Ignore the input text
(b) Focus on relevant parts of the input sequence when generating output
(c) Reduce the number of parameters
(d) Train faster by skipping words

(vi) What is the purpose of data normalization in deep autoencoder?
(a) To scale the input data to a fixed range
(b) To improve the convergence of the optimization algorithm
(c) To preprocess the data for better visualization
(d) To improve the comprehensibility of the input data

(vii) Which of the following metric is commonly used to evaluate the classification performance of a deep autoencoder?
(a) Mean absolute error (b) Mean squared error
(c) Accuracy (d) R-squared

(viii) In transfer learning, the lower layers of a pre-trained model are reused because they:
(a) Contain general-purpose feature representations
(b) Are specific to the target dataset
(c) Contain random noise
(d) Always increase model bias

(ix) Which of the following deep learning techniques is used for unsupervised feature learning?
(a) Convolutional neural networks (b) Recurrent neural networks
(c) Autoencoders (d) Reinforcement learning

(x) 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

Fill in the blanks with the correct word

(xi) In neural networks, learning with a teacher is also referred to as _____ learning.

(xii) The main goal of learning a deep Boltzmann machine is to _____ the overall energy of the network.

(xiii) The father of deep learning is _____.

(xiv) RNNs are mainly used for _____ or _____ data.

(xv) A local minimum in the energy landscape of a Hopfield network is called a _____ state.

Group - B

2. (a) Explain the single layer neural network for classification using the back propagation algorithm for a data set D consisting of the training tuples and their associated target values. *[(CO2)(Apply/IOCQ)]*

(b) Explain why can't the XOR-problem be solved by a one-layer Perceptron? *[(CO2)(Apply/IOCQ)]*

8 + 4 = 12

3. (a) Describe unsupervised, supervised, semi-supervised, and self-supervised learning approaches along with corresponding applications. *[(CO1)(Remember/LOCQ)]*
   (b) Define an independent and identically distributed (i.i.d.) data. *[(CO1)(Remember/LOCQ)]*
   (c) What do you mean by a linear classifier? *[(CO1)(Understand/LOCQ)]*
   (d) State and explain the limitations of linear discriminant analysis. *[(CO1)(Understand/LOCQ)]*

4 + 2 + 3 + 3 = 12

Group - C

4. (a) Justify how a deep convolutional neural network (CNN) can identify an object even when it is transformed. *[(CO4)(Analyse/HOCQ)]*
   (b) Explain the architecture of AlexNet deep convolutional network. *[(CO4)(Analyse/HOCQ)]*

4 + 8 = 12

5. (a) What are the building blocks of a convolutional neural network and how are they used to extract spatial features from input data? *[(CO4)(Apply/IOCQ)]*
   (b) Discuss the working principle of recurrent neural networks and explain how they can be used to process sequential data. *[(CO4)(Apply/IOCQ)]*
   (c) What is the working principle of LSTM networks and how are they used in deep learning applications? *[(CO4)(Apply/IOCQ)]*

4 + 4 + 4 = 12

Group - D

6. (a) How do generative models differ from discriminative models and what are the applications of generative modelling in deep learning? *[(CO4)(Analyse/HOCQ)]*
   (b) Define data-dependent expectation and data-independent expectation in deep Boltzmann machines. *[(CO4)(Apply/IOCQ)]*
   (c) What is the working principle of generative adversarial networks and how are they used for image generation tasks? *[(CO4)(Apply/IOCQ)]*

4 + 4 + 4 = 12

7. (a) What is the working principle of variational autoencoders and how they are used in generative modelling? Mention one limitation of autoencoders. *[(CO4)(Analyse/HOCQ)]*
   (b) Mention the approaches used to estimate the data-dependent expectation and data-independent expectation in deep Boltzmann machines. *[(CO4)(Remember/LOCQ)]*
   (c) Explain greedy layerwise pre-training in context of deep variational autoencoders. *[(CO4)(Apply/IOCQ)]*

(3 + 1) + 4 + 4 = 12

Group - E

8. (a) What is object detection? *[(CO6)(Understand/LOCQ)]*
   (b) Explain which deep architecture is best suited for object detection applications. *[(CO6)(Analyse/HOCQ)]*

(c) Explain how object detection is different from image segmentation.

[(CO6)(Apply/IOCQ)]

4 + 4 + 4 = 12

9. (a) If dropout rate = 0.3 and hidden units = 100, on average, how many hidden units are active during training of an LSTM? *[(CO5)(Apply/IOCQ)]*

(b) If an LSTM layer with input size = 8 and hidden size = 16 has 4 gates, and a GRU has 3 gates, compute the approximate parameter reduction percentage when switching from LSTM to GRU. *[(CO5)(Apply/IOCQ)]*

(c) If a Hopfield network has 100 neurons, estimate the maximum number of patterns it can store reliably. (Hint: use approximate formula 0.15×N).

[(CO5)(Apply/IOCQ)]

3 + 6 + 3 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	20.83	54.17	25