

1. Define Artificial Neuron with equation.

An artificial neuron is a mathematical model inspired by biological neurons.

$$y = f(\sum w_i x_i + b)$$

Where:

- x_i = inputs
 - w_i = weights
 - b = bias
 - f = activation function
-

2. What is the limitation of Perceptron?

- Works only for linearly separable data
 - Cannot solve XOR problem
 - No hidden layer
-

3. Define Backpropagation.

Backpropagation is a supervised learning algorithm used to train multilayer neural networks by computing gradients using the chain rule and updating weights using gradient descent.

4. What is vanishing gradient problem?

During backpropagation, gradients become extremely small in deep networks, preventing early layers from learning effectively. Common in sigmoid/tanh activations.

5. Differentiate between Batch GD and SGD.

Batch GD	SGD
Uses full dataset	Uses single sample
Stable	Noisy
Slow	Fast

6. What is L2 regularization?

Adds penalty term:

$$Loss + \lambda || w ||^2$$

Helps reduce overfitting.

7. Define Convolution in CNN.

Convolution is a mathematical operation where a filter slides over input data to extract spatial features.

8. Role of Pooling Layer?

- Reduces spatial dimension
 - Reduces computation
 - Prevents overfitting
-

9. Define Attention mechanism.

Attention allows model to focus on important parts of input sequence.

$$Attention(Q, K, V) = Softmax(QK^T)V$$

10. What is GAN?

GAN is a generative model consisting of:

- Generator
 - Discriminator
- Trained using adversarial learning.
-

✔ SECTION B (Any 5 × 6 = 30 Marks)

11. Explain Multilayer Perceptron (MLP)

MLP consists of:

- Input layer
- One or more hidden layers
- Output layer

Uses nonlinear activation functions.

Can solve non-linear problems like XOR.

12. Backpropagation Algorithm

Steps:

1. Forward pass
2. Compute loss
3. Compute gradient using chain rule
4. Update weights:

$$w = w - \eta \frac{\partial L}{\partial w}$$

Efficient method for training deep networks.

13. Compare SGD, RMSProp and Adam

Optimizer Feature

SGD Simple update

RMSProp Adaptive learning rate

Adam Momentum + RMSProp

Adam is most widely used.

14. Explain CNN Architecture

Layers:

1. Convolution
2. ReLU
3. Pooling
4. Fully Connected
5. Softmax

Used for image classification.

15. LSTM Network

Contains:

- Forget gate
- Input gate
- Output gate
- Cell state

Solves long-term dependency problem.

16. Transformer Architecture

Main components:

- Multi-head attention
- Feed-forward layer
- Positional encoding
- Layer normalization

Does not use recurrence.

17. Autoencoder & VAE

Autoencoder:

Encoder → Bottleneck → Decoder

VAE:

- Probabilistic latent space
 - Generates new data
-

18. BERT vs GPT

BERT

GPT

Encoder-based Decoder-based

Bidirectional Left-to-right

Classification Generation

 **SECTION C (Any 2 × 10 = 20 Marks)**

19. (a) XOR Problem and MLP

XOR truth table is not linearly separable.

Perceptron cannot solve it.

MLP with hidden layer creates non-linear decision boundary.

(b) Heuristics for improving Backprop:

- Proper learning rate
 - Weight initialization
 - Batch normalization
 - Momentum
 - Adam optimizer
-

20. (a) ResNet Architecture

- Uses skip connections:

$$y = F(x) + x$$

Prevents vanishing gradient.

(b) Transfer Learning

- Use pretrained model (ImageNet)
- Freeze early layers
- Fine-tune final layers

Advantages:

- Less data needed
 - Faster training
-

21. (a) Multi-head Attention

Multiple attention layers run in parallel.

$$Attention(Q, K, V) = Softmax(QK^T / \sqrt{d_k})V$$

Improves learning capacity.

(b) GAN Architecture

Generator produces fake data.

Discriminator distinguishes real/fake.

Minimax objective:

$$\min_G \max_D V(D, G)$$

Applications:

- Image generation
- Deepfake
- Super-resolution

 **Total: 70 Marks (Solved)**

✔ **UNIT 1: Neural Networks & Backpropagation (15 MCQs)**

1. The artificial neuron is inspired by:
A) CPU
B) **Biological neuron**
C) GPU
D) Transistor
2. The output of a perceptron is:
A) Continuous
B) **Binary**
C) Multi-class
D) Random
3. XOR problem is:
A) Linearly separable
B) Regression
C) **Non-linearly separable**
D) Clustering
4. Backpropagation uses:
A) Bayes theorem
B) PCA
C) **Chain rule**
D) KNN
5. Activation function introduces:
A) Linearity
B) **Non-linearity**
C) Noise
D) Bias
6. ReLU function is:
A) x^2
B) $1/x$
C) **$\max(0,x)$**
D) \tanh

7. Sigmoid output range is:
- A) $(-1,1)$
 - B) $(0,\infty)$
 - C) **$(0,1)$**
 - D) $(-\infty,\infty)$
8. MLP stands for:
- A) Multi Linear Perceptron
 - B) **Multilayer Perceptron**
 - C) Multi Logistic Processor
 - D) None
9. Bias term shifts:
- A) Loss
 - B) **Decision boundary**
 - C) Gradient
 - D) Data
10. Learning rate controls:
- A) Accuracy
 - B) Data size
 - C) **Step size of update**
 - D) Epoch number
11. Vanishing gradient occurs in:
- A) CNN
 - B) **Deep networks**
 - C) GAN
 - D) SVM
12. Tanh range is:
- A) $(0,1)$
 - B) **$(-1,1)$**
 - C) $(0,\infty)$
 - D) $(-\infty,\infty)$
13. Backpropagation minimizes:
- A) Accuracy
 - B) **Loss function**
 - C) Input size
 - D) Epoch

14. Perceptron cannot solve:

- A) AND
- B) OR
- C) **XOR**
- D) NAND

15. Hidden layer enables:

- A) Linear mapping
 - B) **Non-linear mapping**
 - C) Clustering
 - D) Sorting
-

 **UNIT 2: Optimization & Regularization (15 MCQs)**

1. Gradient Descent updates weights using:

- A) Random search
- B) **Gradient of loss**
- C) KNN
- D) PCA

2. SGD updates using:

- A) Full dataset
- B) **Single sample**
- C) No sample
- D) Entire batch

3. Momentum helps in:

- A) Overfitting
- B) **Reducing oscillations**
- C) Clustering
- D) Sorting

4. RMSProp adjusts:

- A) Bias
- B) **Learning rate adaptively**
- C) Data size
- D) Loss

5. Adam combines:
 - A) CNN + RNN
 - B) **Momentum + RMSProp**
 - C) SGD + PCA
 - D) GAN + CNN

6. L2 regularization reduces:
 - A) Bias
 - B) **Overfitting**
 - C) Underfitting
 - D) Data

7. Dropout works by:
 - A) Adding neurons
 - B) **Removing neurons randomly**
 - C) Increasing data
 - D) Reducing data

8. Early stopping prevents:
 - A) Underfitting
 - B) **Overfitting**
 - C) Noise
 - D) Clustering

9. Mini-batch GD uses:
 - A) One sample
 - B) **Subset of data**
 - C) Full data
 - D) No data

10. Regularization adds penalty to:
 - A) Input
 - B) Output
 - C) **Weights**
 - D) Labels

11. Dataset augmentation increases:
 - A) Loss
 - B) **Data diversity**
 - C) Bias
 - D) Gradient

12. Semi-supervised learning uses:
- A) Only labeled
 - B) Only unlabeled
 - C) **Both labeled & unlabeled**
 - D) None
13. Adam optimizer is good for:
- A) Slow convergence
 - B) **Fast convergence**
 - C) No training
 - D) Random search
14. Weight decay is another name for:
- A) Dropout
 - B) **L2 regularization**
 - C) SGD
 - D) RMSProp
15. Learning rate too high causes:
- A) Slow learning
 - B) **Divergence**
 - C) Convergence
 - D) Accuracy
-

 **UNIT 3: CNN (15 MCQs)**

1. CNN is mainly used for:
- A) Text
 - B) **Images**
 - C) Audio only
 - D) Sorting
2. Convolution extracts:
- A) Labels
 - B) **Features**
 - C) Bias
 - D) Gradient

3. Filter is also called:
 - A) Pool
 - B) **Kernel**
 - C) Bias
 - D) Node

4. Pooling reduces:
 - A) Channels
 - B) **Spatial dimension**
 - C) Accuracy
 - D) Loss

5. Max pooling selects:
 - A) Average value
 - B) Minimum
 - C) **Maximum value**
 - D) Random

6. Padding helps maintain:
 - A) Loss
 - B) **Output size**
 - C) Filter size
 - D) Bias

7. Stride controls:
 - A) Filter size
 - B) **Step movement**
 - C) Output layer
 - D) Epoch

8. ReLU in CNN adds:
 - A) Linearity
 - B) **Non-linearity**
 - C) Noise
 - D) Bias

9. LeNet was developed for:
 - A) Speech
 - B) **Digit recognition**
 - C) GAN
 - D) NLP

10. AlexNet introduced:
- A) LSTM
 - B) **ReLU & Dropout**
 - C) GAN
 - D) Attention
11. VGG uses:
- A) 5x5 filters
 - B) **3x3 filters**
 - C) RNN
 - D) GAN
12. ResNet uses:
- A) Pooling
 - B) **Skip connections**
 - C) LSTM
 - D) SVM
13. EfficientNet scales:
- A) Only depth
 - B) Only width
 - C) **Depth, width, resolution**
 - D) None
14. Transfer learning uses:
- A) Random weights
 - B) **Pretrained model**
 - C) No weights
 - D) GAN
15. Fully connected layer performs:
- A) Feature extraction
 - B) **Classification**
 - C) Pooling
 - D) Padding
-

✔ **UNIT 4: RNN & LSTM (15 MCQs)**

1. RNN is used for:
 - A) Images
 - B) **Sequential data**
 - C) Clustering
 - D) Sorting
2. RNN maintains:
 - A) Memory
 - B) **Hidden state**
 - C) Bias
 - D) Kernel
3. Vanishing gradient affects:
 - A) Output layer
 - B) **Early layers**
 - C) Pooling
 - D) Labels
4. LSTM stands for:
 - A) Linear State Machine
 - B) **Long Short-Term Memory**
 - C) Large System Model
 - D) None
5. LSTM has how many gates?
 - A) 1
 - B) 2
 - C) **3**
 - D) 4
6. GRU has:
 - A) 3 gates
 - B) **2 gates**
 - C) 4 gates
 - D) 5 gates

7. Bidirectional RNN reads:
 - A) Forward only
 - B) Backward only
 - C) **Both directions**
 - D) None

8. Time series forecasting uses:
 - A) CNN
 - B) **RNN/LSTM**
 - C) GAN
 - D) SVM

9. Forget gate controls:
 - A) Output
 - B) **Memory retention**
 - C) Bias
 - D) Filter

10. GRU is:
 - A) Complex than LSTM
 - B) **Simpler than LSTM**
 - C) Same
 - D) None

11. RNN weight sharing occurs across:
 - A) Layers
 - B) **Time steps**
 - C) Filters
 - D) Nodes

12. Activation in RNN commonly:
 - A) Softmax
 - B) **Tanh**
 - C) Linear
 - D) None

13. LSTM solves:
 - A) Overfitting
 - B) **Long-term dependency**
 - C) Clustering
 - D) Sorting

14. Sentiment analysis uses:

- A) GAN
- B) **LSTM**
- C) PCA
- D) KNN

15. Sequence length affects:

- A) Bias
 - B) **Gradient stability**
 - C) Filter
 - D) Loss
-

 **UNIT 5: Attention & Transformers (15 MCQs)**

1. Transformer removes:

- A) CNN
- B) **RNN**
- C) GAN
- D) Pooling

2. Attention uses:

- A) Only query
- B) **Query, Key, Value**
- C) Loss
- D) Bias

3. Self-attention operates on:

- A) Multiple datasets
- B) **Same sequence**
- C) Images only
- D) None

4. Multi-head attention improves:

- A) Speed
- B) **Representation learning**
- C) Loss
- D) Bias

5. Positional encoding provides:
 - A) Loss
 - B) **Order information**
 - C) Bias
 - D) Kernel

6. BERT is:
 - A) Decoder
 - B) **Encoder-based**
 - C) GAN
 - D) CNN

7. GPT is:
 - A) Encoder
 - B) **Decoder-based**
 - C) CNN
 - D) GAN

8. Transformer uses:
 - A) Recurrence
 - B) **Parallel processing**
 - C) Loop
 - D) None

9. Autoencoder consists of:
 - A) RNN
 - B) **Encoder & Decoder**
 - C) GAN
 - D) CNN

10. VAE is used for:
 - A) Classification
 - B) **Generation**
 - C) Sorting
 - D) Clustering

11. Sparse autoencoder enforces:
 - A) Noise
 - B) **Sparsity constraint**
 - C) Bias
 - D) Kernel

12. Denoising AE removes:
- A) Bias
 - B) **Noise**
 - C) Filter
 - D) Loss
13. Attention score computed by:
- A) Addition
 - B) **Dot product**
 - C) Division
 - D) None
14. BERT training objective:
- A) Regression
 - B) **Masked language modeling**
 - C) Clustering
 - D) Sorting
15. GPT is mainly for:
- A) Classification
 - B) **Text generation**
 - C) Clustering
 - D) Filtering
-

 **UNIT 6: GAN (15 MCQs)**

1. GAN has:
- A) 1 network
 - B) **2 networks**
 - C) 3 networks
 - D) 4 networks
2. Generator creates:
- A) Labels
 - B) **Fake data**
 - C) Loss
 - D) Gradient

3. Discriminator acts as:
 - A) Generator
 - B) **Classifier (real/fake)**
 - C) Encoder
 - D) Decoder

4. GAN training is:
 - A) Cooperative
 - B) **Adversarial**
 - C) Supervised
 - D) Unsupervised only

5. DCGAN uses:
 - A) RNN
 - B) **CNN**
 - C) LSTM
 - D) PCA

6. CycleGAN performs:
 - A) Classification
 - B) **Image translation**
 - C) Sorting
 - D) Clustering

7. StyleGAN is famous for:
 - A) Text
 - B) **High-quality faces**
 - C) Audio
 - D) Sorting

8. GAN objective is:
 - A) Minimize loss only
 - B) **Minimax game**
 - C) Regression
 - D) Clustering

9. GAN can generate:
 - A) Images
 - B) Audio
 - C) Text
 - D) **All of the above**

10. Mode collapse occurs when:
- A) Good diversity
 - B) **Generator produces similar outputs**
 - C) Stable training
 - D) None
11. GAN training is often:
- A) Easy
 - B) **Unstable**
 - C) Deterministic
 - D) Fixed
12. Latent vector is:
- A) Output
 - B) **Input noise**
 - C) Label
 - D) Loss
13. GAN is used in:
- A) Deepfake
 - B) Super-resolution
 - C) Image synthesis
 - D) **All of the above**
14. Discriminator output is usually:
- A) Multi-class
 - B) **Binary**
 - C) Regression
 - D) None
15. Generator tries to:
- A) Increase loss
 - B) **Fool discriminator**
 - C) Classify data
 - D) Reduce input