# 📘 Neural Networks & Deep Learning (Module I) – Interview Q&A

## ****Section 1: Basics of Neural Networks****

**Q1. What is a neuron in a neural network?**  
A neuron is like a small calculator. It multiplies input by weights, adds bias, passes through an activation function, and produces output.  
👉 Example: y = sigmoid(w·x + b)

**Q2. What are weights in a neural network?**  
Weights are numbers that show how important an input is. The network learns them during training.  
👉 Example: In image recognition, weights detect edges.

**Q3. What is bias in a neural network?**  
Bias is like an extra input that helps shift the output. It makes the model flexible.  
👉 Example: y = w·x + b

**Q4. What is variance?**  
Variance means how much the model’s predictions change with different data.  
👉 High variance = overfitting (model memorizes instead of generalizing).

**Q5. What is underfitting?**  
Underfitting = model too simple, fails to learn.  
👉 Example: Using a straight line for curved data.

**Q6. What is overfitting?**  
Overfitting = model learns training data too well, fails on new data.  
👉 Example: Student memorizing answers without understanding.

**Q7. What is generalization?**  
Generalization = ability of a model to perform well on unseen data.  
👉 Example: A spam filter that works on new emails.

## ****Section 2: Activation Functions****

**Q8. What is an activation function?**  
It adds non-linearity so the model can learn complex patterns. Without it, NN is just linear regression.

**Q9. Sigmoid function?**  
Maps values between 0 and 1 → good for probabilities.  
👉 Problem: vanishing gradient.

**Q10. Tanh function?**  
Maps values between -1 and 1. Zero-centered.  
👉 Better than sigmoid in many cases.

**Q11. ReLU function?**  
ReLU = max(0, x).  
👉 Fast, widely used, but neurons can “die” if stuck at 0.

**Q12. Leaky ReLU?**  
Leaky ReLU allows small negative values (e.g., 0.01x).  
👉 Solves dying ReLU problem.

**Q13. Softmax function?**  
Converts numbers into probabilities that sum to 1.  
👉 Example: Used in multi-class classification.

## ****Section 3: Training & Loss Functions****

**Q14. What is a loss function?**  
Measures error between prediction and true value.  
👉 Example:

MSE (regression)

Cross-Entropy (classification)

**Q15. What is forward propagation?**  
It’s passing inputs through the network to get predictions.

**Q16. What is backpropagation?**  
It’s how the network learns: calculate error, use chain rule, update weights.

**Q17. What is gradient descent?**  
Optimization method to reduce loss by updating weights step by step.

**Q18. Batch Gradient Descent?**  
Uses whole dataset for update → stable but slow.

**Q19. Stochastic Gradient Descent (SGD)?**  
Uses 1 sample at a time → faster but noisy.

**Q20. Mini-batch Gradient Descent?**  
Uses small groups of samples → balance of speed & stability.

## ****Section 4: Optimizers****

**Q21. What is Momentum optimizer?**  
Adds part of previous update → speeds up learning.  
👉 Like rolling a ball downhill.

**Q22. What is RMSProp optimizer?**  
Adjusts learning rate for each weight using moving averages.  
👉 Good for RNNs.

**Q23. What is Adam optimizer?**  
Combines Momentum + RMSProp.  
👉 Most widely used, default choice in deep learning.

**Q24. What is Adagrad optimizer?**  
Adapts learning rate → larger updates for rare features.  
👉 Useful for text data.

**Q25. What is learning rate?**  
Step size for weight updates.  
👉 Too high = unstable, too low = slow.

## ****Section 5: Regularization & Generalization****

**Q26. What is dropout?**  
Randomly ignores some neurons during training → prevents overfitting.

**Q27. What is regularization?**  
Adds penalty to weights → prevents overfitting.  
👉 Types: L1, L2.

**Q28. What is L1 regularization?**  
Adds sum of absolute weights.  
👉 Creates sparse models (many weights = 0).

**Q29. What is L2 regularization?**  
Adds sum of squared weights.  
👉 Prevents large weights.

**Q30. What is Early Stopping?**  
Stop training when validation error increases → avoids overfitting.

## ****Section 6: Neural Network Architecture****

**Q31. What is a perceptron?**  
A single-layer NN → works for simple linear problems.

**Q32. What is Multi-Layer Perceptron (MLP)?**  
NN with hidden layers → learns complex patterns.

**Q33. What is deep learning?**  
Subset of ML → uses deep neural networks with many layers.

**Q34. Difference between AI, ML, and DL?**

AI = broad field (machines acting smart)

ML = learning from data

DL = uses neural networks

**Q35. Why use hidden layers?**  
They extract features step by step.  
👉 Example: Edges → shapes → objects.

**Q36. How to choose number of hidden layers?**  
Trial-and-error. Start small, add layers until accuracy improves.

**Q37. What is vanishing gradient problem?**  
Gradients shrink, network stops learning.  
👉 Happens in deep nets with sigmoid/tanh.

**Q38. What is exploding gradient problem?**  
Gradients become too large → unstable training.  
👉 Solved by gradient clipping.

Q39. What is weight initialization?  
Starting weights smartly to avoid vanishing/exploding.  
👉 Example: Xavier initialization.

**Q40. What is hyperparameter?**  
Values set before training.  
👉 Example: learning rate, batch size, number of layers.

## ****Section 7: Model Evaluati**on**

**Q41. What is cross-validation?**  
Splits data into multiple parts for training/testing.  
👉 Example: K-fold CV.

**Q42. What is bias-variance tradeoff?**  
Balance between underfitting (high bias) and overfitting (high variance).

**Q43. What is epoch?**  
One full pass over training data.

**Q44. What is batch size?**  
Number of samples processed before weights update.

**Q45. What is training vs testing data?**

Training data = learn patterns

Testing data = check performance

**Q46. What is validation data?**  
Used for tuning hyperparameters.

**Q47. What is normalization?**  
Scaling data to range [0,1].  
👉 Example: pixel values ÷ 255.

**Q48. What is standardization?**  
Scaling data to mean = 0, std = 1.

**Q49. What is one-hot encoding?**  
Convert categories into 0/1 vectors.  
👉 Example: Cat=[1,0,0], Dog=[0,1,0].

**Q50. What is cost function?**  
Measures error of model. Same as loss function in many cases.

## ****Section 8: Practical Deep Learning****

**Q51. Why use ReLU instead of sigmoid?**  
ReLU is faster and avoids vanishing gradients.

**Q52. Why is Softmax used?**  
For multi-class classification → gives probabilities.

**Q53. What is feature extraction?**  
Learning important patterns automatically.  
👉 CNN learns edges → shapes → objects.

**Q54. What is chain rule in backpropagation?**  
Used to calculate gradients layer by layer.

**Q55. What is forward vs backward pass?**

Forward pass = calculate output

Backward pass = update weights

**Q56. What is data augmentation?**  
Making more training data by transformations.  
👉 Example: rotating images.

**Q57. What is model capacity?**  
How complex patterns a model can learn.  
👉 Small NN = low capacity, Deep NN = high capacity.

**Q58. What is transfer learning?**  
Using pre-trained models for new tasks.  
👉 Example: ResNet trained on ImageNet reused for medical images.

**Q59. What is representation learning?**  
Learning useful data features automatically.  
👉 Example: Word2Vec learns word meanings.

**Q60. Summarize NN training process.**

Input data

Forward pass (prediction)

Calculate loss

Backpropagation (update weights)

Repeat for many epochs