Neural Network Viva Questions & Answers (Updated)

1. Q: What is a Perceptron?

A: A perceptron is the simplest neural network model that takes inputs, applies weights, adds bias, and passes through an activation function to produce output.

2. Q: Why do we need a bias term in neural networks?

A: Bias allows shifting the activation function, improving flexibility of decision boundaries.

3. Q: What is the difference between Linear Regression and Perceptron?

A: Linear regression outputs continuous values, perceptron classifies with threshold.

4. Q: What is an activation function?

A: Introduces non-linearity in the model. Example: Sigmoid $\sigma(x)=1/(1+e^{-x})$.

5. Q: What is forward propagation?

A: Process of passing inputs through network to compute predictions.

6. Q: What is backpropagation?

A: Weights updated using gradient of loss w.r.t. weights via chain rule.

7. Q: What is the role of learning rate?

A: Controls step size in weight updates. High $\rightarrow$ overshoot, Low $\rightarrow$ slow training.

8. Q: What is a loss function?

A: Measures difference between predicted & actual values. Example: MSE = avg((y_true - y_pred)^2).

9. Q: Give an example of classification loss function.

A: Cross-Entropy Loss. Example: True=1, Pred=0.9 $\rightarrow$ Loss= -log(0.9)=0.105.

10. Q: What is an optimizer?

A: Algorithm used to minimize loss by updating weights. Example: SGD, Adam.

11. Q: What is bias in machine learning?

A: Error due to oversimplification. Example: Straight line fitting curved data.

12. Q: What is variance?

A: Error due to sensitivity to training data. Example: Overfitted tree good on train, poor on test.

13. Q: Explain bias-variance tradeoff.

A: More complexity $\rightarrow$ less bias but more variance.

14. Q: How to reduce overfitting?

A: Use regularization, dropout, early stopping, cross-validation.

15. Q: What is underfitting?

A: Model too simple to capture data. Example: Linear model on quadratic data.

16. Q: What is cross-validation?

A: Evaluating performance by splitting dataset into multiple train-test splits. Example: 5-fold CV.

17. Q: Why use cross-validation?

A: To avoid overfitting and estimate performance reliably.

18. Q: What is confusion matrix?

A: Table showing true vs predicted. Example: [[50,10],[5,35]].

19. Q: Define Precision and Recall.

A: Precision=TP/(TP+FP), Recall=TP/(TP+FN).

20. Q: What is F1 score?

A: Harmonic mean of precision & recall. Example: P=0.8, R=0.6 → F1=0.685.

21. Q: What is Gradient Descent?

A: Optimization method to minimize loss by moving opposite to gradient.

22. Q: Difference between SGD and Batch Gradient Descent?

A: SGD updates per sample, Batch after full dataset.

23. Q: Why is Adam optimizer popular?

A: Combines momentum & adaptive learning rate for faster convergence.

24. Q: What is Momentum?

A: Accelerates updates using past gradients. Like a rolling ball.

25. Q: What is RMSProp?

A: Uses moving average of squared gradients for adaptive learning rate.

26. Q: What is a feedforward neural network?

A: Info flows one-way (input→hidden→output).

27. Q: What is a CNN?

A: Specialized for image tasks using convolution & pooling layers.

28. Q: What is an RNN?

A: Has feedback loops, good for sequential data. Example: next-word prediction.

29. Q: What is LSTM?

A: RNN variant with memory cells to solve vanishing gradient problem.

30. Q: What is dropout?

A: Randomly drops neurons during training to prevent overfitting.

31. Q: What is vanishing gradient problem?

A: Gradients shrink to zero in deep nets, slowing learning.

32. Q: What is exploding gradient problem?

A: Gradients become too large, causing instability.

33. Q: What is batch normalization?

A: Normalizes inputs of each layer for stable training.

34. Q: Difference between supervised and unsupervised learning?

A: Supervised uses labeled data, unsupervised uses unlabeled.

35. Q: What is reinforcement learning?

A: Learning with trial/error and rewards. Example: AI in chess.

36. Q: Train a perceptron on AND gate.

A: Inputs (0,0),(0,1),(1,0),(1,1). Outputs (0,0,0,1). Learned weights classify correctly.

37. Q: Explain bias & variance with example.

A: House price prediction: High bias $\rightarrow$ same prediction always. High variance $\rightarrow$ memorizes training noise.

38. Q: Why need non-linear activation?

A: Without it, network is just linear. Example: ReLU allows complex mapping.

39. Q: What is early stopping?

A: Stop when validation error rises though training error decreases.

40. Q: Why use cross-entropy loss in classification?

A: Measures distance between predicted vs true distribution.

41. Q: What is regularization?

A: Techniques to reduce overfitting by adding constraints.

42. Q: What is L1 regularization?

A: Adds absolute weights. Encourages sparsity.

43. Q: What is L2 regularization?

A: Adds squared weights. Prevents large weights.

44. Q: What is Elastic Net?

A: Combination of L1 and L2.

45. Q: How does dropout act as regularization?

A: Randomly drops neurons (output=0) during training.

46. Q: What is Early Stopping?

A: Stop when validation loss increases.

47. Q: What is Data Augmentation?

A: Expanding dataset with transformations.

48. Q: How does weight decay relate to L2?

A: Weight decay is another name for L2 regularization.

49. Q: Example where regularization is necessary?

A: Overfitting small dataset → apply L2 + dropout.

50. Q: How to choose λ in regularization?

A: Use cross-validation. Too high λ=underfit, too low=overfit.

51. Q: What is the role of an activation function?

A: It introduces non-linearity into neural networks, enabling learning of complex patterns.

52. Q: Explain Sigmoid activation function.

A: $\sigma(x)=1/(1+e^{-x})$. Outputs between 0 and 1. Good for probability but suffers from vanishing gradient.

53. Q: Explain Tanh activation function.

A: Outputs between -1 and 1. Zero-centered, better than sigmoid but still vanishes.

54. Q: What is ReLU?

A: f(x)=max(0,x). Fast, avoids vanishing gradient. Problem: Dying ReLU.

55. Q: What is Leaky ReLU?

A: Allows small slope for negative values. Solves dying ReLU.

56. Q: What is Softmax?

A: Converts outputs into probability distribution. Example: input [2,1,0] → [0.71,0.26,0.03].

57. Q: Which activation is best for hidden layers?

A: ReLU or its variants are widely used.

58. Q: Why not use sigmoid in deep hidden layers?

A: Because it causes vanishing gradients.

59. Q: Difference between optimizer & loss function?

A: Loss = error measure, Optimizer = algorithm to reduce it.

60. Q: Explain SGD.

A: Updates weights per sample. Fast but noisy.

61. Q: What is Mini-Batch GD?

A: Updates using small subset of data. Balance between Batch & SGD.

62. Q: Explain Adam optimizer.

A: Combines momentum and adaptive learning rates. Faster convergence.

63. Q: Compare SGD vs Adam.

A: SGD simple but needs tuning; Adam adaptive and faster.

64. Q: What is Momentum?

A: Adds fraction of past gradient to speed convergence.

65. Q: Difference between RMSProp & Adam?

A: RMSProp uses squared gradients, Adam adds momentum too.

66. Q: Explain L1 regularization in detail.

A: L1 adds sum of absolute weights to loss. Encourages sparsity, useful in feature selection.

67. Q: Explain L2 regularization in detail.

A: L2 adds sum of squared weights to loss. Keeps weights small, improves generalization.

68. Q: What is a learning curve?

A: A plot of model error vs training epochs. Helps detect underfitting/overfitting.

69. Q: What is transfer learning?

A: Reusing pre-trained models on new tasks. Example: Using ResNet trained on ImageNet for medical images.

70. Q: What is gradient clipping?

A: Technique to prevent exploding gradients by capping their values.

71. Q: What is one-hot encoding?

A: Method to represent categorical variables as binary vectors. Example: Cat, Dog $\rightarrow$ [1,0], [0,1].

72. Q: What is the difference between epoch, batch, and iteration?

A: Epoch=one pass through dataset. Batch=subset of data. Iteration=one weight update step.