DEEP LEARNING BASICS & ADVANCED

NEURAL NETWORKS FUNDAMENTALS

1. What is a Neural Network? How is it inspired by biological neurons?
2. What is an Artificial Neural Network (ANN)?
3. What is a perceptron? Explain single layer perceptron.
4. What is a Multi-layer Perceptron (MLP)?
5. How is Deep Learning different from traditional ML?
6. When should you use Deep Learning?
7. What are input layer, hidden layer, and output layer?
8. What is a neuron / node in a neural network?
9. What are weights and biases?
10. What is forward propagation? Explain step by step.

Activation Functions

11. What is an activation function? Why is it necessary?
12. What is the problem with linear activation function?
13. What is the sigmoid activation function? Formula and graph.
14. What are the limitations of the sigmoid function?
15. What is the tanh (Hyperbolic Tangent) activation function?
16. Tanh vs Sigmoid - which is better and why?
17. What is ReLU (Rectified Linear Unit)? State the formula.
18. Why is ReLU popular? What are its advantages?
19. What is the dying ReLU problem?
20. What is Leaky ReLU? How does it solve the problem?
21. What is Parametric ReLU (PReLU)?
22. What is ELU (Exponential Linear Unit)?
23. What is the Swish activation function?
24. What is the softmax activation function? When is it used?
25. Which activation function should you use in the output layer (regression vs classification)?

BACKPROPAGATION & TRAINING

Training Neural Networks

26. What is the backpropagation algorithm? Explain in detail.
27. How is the chain rule of calculus used in backpropagation?
28. How does gradient descent work in neural networks?
29. What is a loss function / cost function?
30. When should you use Mean Squared Error (MSE) loss?
31. What is cross-entropy loss? Why is it used in classification?
32. What is the difference between binary cross-entropy and categorical cross-entropy?
33. What is an epoch in training?
34. What is a batch? What does batch size mean?
35. What is an iteration?

Optimization Algorithms

36. What is learning rate? How do you choose the optimal value?
37. What problems arise if the learning rate is too high or too low?
38. What is Batch Gradient Descent?
39. What is Stochastic Gradient Descent (SGD)?
40. What is Mini-batch Gradient Descent?
41. What is momentum in optimization?
42. How does SGD with Momentum work?
43. What is the AdaGrad optimizer?
44. What is the RMSprop optimizer?
45. What is the Adam optimizer? Why is it popular?

REGULARIZATION & OPTIMIZATION

Regularization Techniques

46. How does overfitting occur in neural networks?
47. How do you use L1 and L2 regularization in neural networks?
48. What is dropout? How does it work?
49. What is the dropout rate typically?
50. How is dropout different in training and inference?
51. What is Batch Normalization? Why do we use it?

52. Where should you apply Batch Normalization (before/after activation)?
53. What is Layer Normalization?
54. What is the difference between Batch Norm and Layer Norm?
55. What is Early Stopping?
56. What is Data Augmentation? Give an example.

Weight Initialization

57. Why is weight initialization important?
58. What is the problem with initializing all weights to zero?
59. Why is random initialization better?
60. What is Xavier/Glorot initialization?
61. What is He initialization? Why is it used with ReLU?

CONVOLUTIONAL NEURAL NETWORKS - CNN

CNN Basics

62. What is a CNN? How is it different from a fully connected network?
63. Why is CNN better for image processing?
64. What is a convolution operation? Give a visual example.
65. What is a filter / kernel in CNN?
66. What is a feature map?
67. What is stride in convolution?
68. What is padding? Valid padding vs Same padding.
69. What is the receptive field?
70. What is parameter sharing in CNN?
71. What is translation invariance?

CNN Architecture Components

72. What does the convolutional layer do?
73. What is a pooling layer? What is its purpose?
74. What is Max Pooling?
75. What is Average Pooling?
76. What is Global Average Pooling?
77. When should you use Max Pooling vs Average Pooling?
78. What role does the Fully Connected layer play in CNN?

79. What is flattening in CNN?
80. What is the typical architecture flow in CNN?

Famous CNN Architectures

81. What is the LeNet architecture? (Historical importance)
82. What was AlexNet? Why did it bring a revolution?
83. What is the specialty of VGGNet architecture?
84. What is ResNet? Explain residual connections.
85. What is the main idea of the Inception (GoogLeNet) architecture?

RECURRENT NEURAL NETWORKS - RNN

RNN Basics

86. What is an RNN? How is it different from a feedforward network?
87. What is sequential data? Give examples.
88. What is the "memory" concept in RNN?
89. What role does the hidden state play in RNN?
90. What is the basic equation of RNN?
91. What is the vanishing gradient problem in RNN?
92. What is the exploding gradient problem?
93. How do you solve the vanishing gradient problem?

Advanced RNN Architectures

94. What is LSTM (Long Short-Term Memory)?
95. What are cell state, forget gate, input gate, and output gate in LSTM?
96. How does LSTM solve the vanishing gradient problem?
97. What is GRU (Gated Recurrent Unit)?
98. LSTM vs GRU - comparison.
99. What is Bidirectional RNN? Give a use case.
100. What are sequence-to-sequence models? Give an example.