Chapter 4 MCQs

1. What serves as the backbone of modern deep learning systems?
    A. Relational databases
    B. Neural networks
    C. Operating systems
    D. Word processors
2. Neural networks are composed of layers of artificial neurons, also known as what?
    A. Circuits
    B. Nodes
    C. Switches
    D. Registers
3. During training, how do neural networks adjust internal weights and biases?
    A. Randomly
    B. Manually
    C. Using a method like backpropagation
    D. Through external programming
4. Based on complexity and the number of hidden layers, neural networks are broadly categorized into which two types?
    A. Simple and complex
    B. Shallow and deep
    C. Linear and non-linear
    D. Static and dynamic
5. How many hidden layers do Shallow Neural Networks typically consist of?
    A. Zero
    B. One or two
    C. More than five
    D. Ten or more
6. Deep Neural Networks (DNNs) are ideal for complex tasks because they can learn what?
    A. Linear relationships
    B. Hierarchical feature representations
    C. Simple patterns only
    D. Fixed rules
7. Which type of neural network architecture is primarily designed for image processing tasks?
    A. Feedforward Neural Networks (FNNs)
    B. Recurrent Neural Networks (RNNs)
    C. Convolutional Neural Networks (CNNs)
    D. Transformers
8. Which neural network architecture is best suited for sequential data like time series and text?
    A. Feedforward Neural Networks (FNNs)
    B. Convolutional Neural Networks (CNNs)
    C. Recurrent Neural Networks (RNNs)
    D. Perceptrons
9. What is an advanced variant of RNNs that specifically handles long-term dependencies?
    A. FNN
    B. CNN

C. LSTM
D. Perceptron

10. Which neural network architecture is widely used in modern Natural Language Processing (NLP) models like BERT and GPT?
    A. LSTMs
    B. CNNs
    C. Transformers
    D. FNNs
11. Who introduced the Perceptron model in 1958?
    A. Geoffrey Hinton
    B. Frank Rosenblatt
    C. Yann LeCun
    D. Andrew Ng
12. What kind of output does a basic Perceptron produce?
    A. Continuous values
    B. Probabilities
    C. Binary outputs (0 or 1)
    D. Categorical labels
13. Which component of a Perceptron tells the importance of a feature in predicting the final value?
    A. Input values
    B. Bias
    C. Weights
    D. Activation function
14. What is the purpose of the bias term in a Perceptron?
    A. To multiply the input values
    B. To shift the activation function
    C. To determine the final output
    D. To sum the weighted inputs
15. What does the summation function in a Perceptron do?
    A. Introduces non-linearity
    B. Binds the weights and inputs together to find their sum
    C. Determines the output threshold
    D. Selects the most important features
16. What is the primary limitation of a single-layer perceptron?
    A. It can only solve non-linear problems.
    B. It requires too much computational power.
    C. It can only solve linearly separable problems.
    D. It struggles with small datasets.
17. Why does a single-layer perceptron have limited expressive power?
    A. It has too many hidden layers.
    B. It cannot learn intermediate abstractions due to no hidden layer.
    C. It uses too many activation functions.
    D. It is computationally expensive.
18. What is a key enhancement of Multi-Layer Perceptrons (MLPs) over single-layer perceptrons?
    A. They have fewer layers.
    B. They can model non-linear decision boundaries.
    C. They use only linear activation functions.
    D. They are simpler to train.

19. What algorithm is used to train MLPs by efficiently computing gradients of the loss function?
    A. Linear Regression
    B. K-Means Clustering
    C. Backpropagation
    D. Decision Tree Induction
20. What theorem states that an MLP with at least one hidden layer can approximate any continuous function?
    A. Central Limit Theorem
    B. No Free Lunch Theorem
    C. Universal Approximation Theorem
    D. Bayes' Theorem
21. What is the main purpose of activation functions in neural networks?
    A. To store data
    B. To introduce non-linearity into the model
    C. To define the network's input size
    D. To reduce the number of neurons
22. Without activation functions, a neural network, regardless of its depth, would behave like what type of model?
    A. A classification model
    B. A linear regression model
    C. A clustering model
    D. A decision tree
23. Which activation function transforms its input into a value between 0 and 1, suitable for probabilities?
    A. ReLU
    B. Tanh
    C. Sigmoid
    D. Softmax
24. What is a major drawback of the Sigmoid activation function, especially in deep networks?
    A. It is computationally expensive.
    B. It suffers from the vanishing gradient problem.
    C. Its output is zero-centered.
    D. It only outputs negative values.
25. Which activation function outputs values between -1 and 1 and is zero-centered?
    A. Sigmoid
    B. ReLU
    C. Tanh
    D. Leaky ReLU
26. What is a common use case for the Sigmoid activation function?
    A. Hidden layers in CNNs
    B. Output layer of binary classification models
    C. Multi-class classification output
    D. Regression tasks
27. Which activation function outputs the input directly if it is positive, and zero otherwise?
    A. Sigmoid
    B. Tanh
    C. ReLU
    D. Softmax

28. What is a major drawback of the ReLU activation function?
    A. It saturates in the positive region.
    B. It suffers from the vanishing gradient problem in the positive region.
    C. It can suffer from the "dying ReLU" problem.
    D. It is computationally inefficient.
29. Which activation function is a variant of ReLU that allows a small, non-zero gradient for negative inputs?
    A. Sigmoid
    B. Tanh
    C. Leaky ReLU
    D. Softmax
30. What is the primary use case for the Softmax activation function?
    A. Binary classification output layer
    B. Hidden layers in deep networks
    C. Output layer of multi-class classification models
    D. Regression tasks
31. For hidden layers in deep neural networks, which activation function is most commonly used due to its efficiency and effectiveness?
    A. Sigmoid
    B. Tanh
    C. ReLU
    D. Softmax
32. If a network suffers from the "dying neuron problem," which activation function might be preferred over ReLU?
    A. Sigmoid
    B. Tanh
    C. Leaky ReLU
    D. Softmax
33. What is a Feedforward Neural Network (FNN) characterized by?
    A. Feedback loops
    B. Data flow in one direction only
    C. Recurrent connections
    D. Cycles in its architecture
34. In an FNN, what is the first layer where raw input data is fed into the model?
    A. Hidden layer
    B. Output layer
    C. Input layer
    D. Processing layer
35. What happens in the hidden layers of an FNN?
    A. Raw input data is received.
    B. The model provides its final predictions.
    C. The network processes the input data and learns complex patterns.
    D. Only linear operations are performed.
36. What do activation functions like ReLU, sigmoid, and tanh introduce into an FNN?
    A. Linearity
    B. Non-linearity
    C. Randomness
    D. Bias
37. In a regression task, how many neurons does the output layer of an FNN typically have?
    A. Multiple neurons

B. A single neuron
C. Zero neurons
D. Depends on the number of classes

38. What are the crucial parameters for an FNN's learning process?
    A. Input data and output data
    B. Layers and neurons
    C. Weights and biases
    D. Activation functions and loss functions
39. What is the goal of adjusting weights and biases during FNN training?
    A. To increase the error
    B. To minimize the error between predicted and actual outputs
    C. To maximize the training time
    D. To make the model more complex
40. What is the first step in each epoch during FNN training?
    A. Error is computed.
    B. Error is propagated backward.
    C. Data is passed through the network (forward pass).
    D. Weights and biases are updated.
41. What is the purpose of the "backward pass" in FNN training?
    A. To feed data into the network
    B. To compute the prediction
    C. To propagate the error backward to update weights and biases
    D. To visualize performance
42. In the forward pass of an FNN, what does each neuron compute?
    A. Only the bias term
    B. A weighted sum of its inputs, followed by an activation function
    C. The loss value
    D. The gradient of the loss function
43. What is Backpropagation primarily used for in artificial neural networks?
    A. To define the network architecture
    B. To preprocess input data
    C. To efficiently compute the gradient of the loss function
    D. To visualize training progress
44. What is the primary goal of training a neural network?
    A. To maximize accuracy
    B. To minimize a loss function
    C. To increase computational cost
    D. To simplify the model
45. How does Backpropagation solve the problem of manually computing adjustments in deep networks?
    A. By using random adjustments
    B. By providing a systematic and efficient method for calculating gradients using the chain rule
    C. By avoiding the need for gradients
    D. By only adjusting the output layer
46. What is the first step in the typical learning process of a neural network?
    A. Loss Computation
    B. Backward Pass
    C. Forward Pass
    D. Weight Update

47. After the forward pass, what is the next step in the learning process?
    A. Backward Pass
    B. Weight Update
    C. Loss Computation
    D. Model Deployment
48. What do the gradients computed during the backward pass indicate?
    A. The overall accuracy of the model
    B. How much each parameter contributed to the error
    C. The optimal learning rate
    D. The number of hidden layers needed
49. What are the gradients used for in the final step of the learning process?
    A. To visualize the network
    B. To update the weights and biases in a direction that reduces the loss
    C. To increase the loss
    D. To change the activation functions
50. Backpropagation lies at the heart of what type of learning in deep neural networks?
    A. Unsupervised learning
    B. Reinforcement learning
    C. Semi-supervised learning
    D. Supervised learning
51. What is the primary purpose of a loss function in deep learning?
    A. To increase model complexity
    B. To quantify the difference between predicted and actual values
    C. To speed up data loading
    D. To visualize network architecture
52. Which type of loss function is typically used for regression problems?
    A. Binary Cross-Entropy
    B. Categorical Cross-Entropy
    C. Mean Squared Error (MSE)
    D. Softmax Loss
53. What characteristic of Mean Squared Error (MSE) makes it sensitive to outliers?
    A. It uses absolute differences.
    B. It penalizes large errors more heavily due to squaring.
    C. It provides non-smooth gradients.
    D. It encourages the model to find the median.
54. Which regression loss function is more robust to outliers?
    A. Mean Squared Error (MSE)
    B. Mean Absolute Error (MAE)
    C. Huber Loss
    D. Root Mean Squared Error (RMSE)
55. What is a drawback of Mean Absolute Error (MAE)?
    A. It is sensitive to outliers.
    B. Its gradients are not smooth at zero.
    C. It encourages the model to find the mean.
    D. It leads to overly smooth predictions.
56. Binary Cross-Entropy Loss is used for what type of problems?
    A. Regression
    B. Multi-class classification
    C. Binary classification
    D. Time-series prediction

57. How does Binary Cross-Entropy Loss penalize incorrect predictions?
    A. Linearly
    B. Quadratically
    C. Logarithmically
    D. Exponentially
58. Categorical Cross-Entropy Loss extends which other loss function to multi-class problems?
    A. Mean Squared Error
    B. Mean Absolute Error
    C. Binary Cross-Entropy
    D. Huber Loss
59. What does Categorical Cross-Entropy Loss encourage the predicted probability for the true class to be?
    A. Close to 0
    B. Close to 0.5
    C. Close to 1
    D. Random
60. What is the main job of an optimizer in a neural network?
    A. To define the network's layers
    B. To minimize the loss function
    C. To preprocess input data
    D. To visualize training progress
61. What does choosing the right optimizer impact?
    A. How fast your model learns
    B. How well it performs
    C. How stable the training is
    D. All of the above
62. Which basic optimizer updates the model after looking at a few samples at a time (a mini-batch)?
    A. Adam
    B. RMSprop
    C. Stochastic Gradient Descent (SG
    D. AdaGrad
63. Which optimizer combines the strengths of both Momentum and RMSprop?
    A. SGD
    B. AdaGrad
    C. Adam
    D. SGD with Momentum
64. Which optimizer is particularly helpful when gradients are inconsistent, such as in Recurrent Neural Networks (RNNs)?
    A. SGD
    B. Adam
    C. RMSprop
    D. AdaGrad
65. Which optimizer is great for sparse data like text because it gives less attention to frequent updates?
    A. Adam
    B. RMSprop
    C. AdaGrad
    D. SGD with Momentum

66. What occurs when a model learns the training data too well, including noise, and performs poorly on unseen data?
    A. Underfitting
    B. Optimal fitting
    C. Overfitting
    D. Generalization
67. What is a classic symptom of overfitting?
    A. Training loss increases while validation loss decreases.
    B. Both training and validation loss remain low.
    C. Training loss continues to decrease while validation loss starts increasing.
    D. Training accuracy and validation accuracy are both high.
68. Which of the following is a strategy to prevent overfitting?
    A. Reducing the amount of training data
    B. Increasing model complexity
    C. Data Augmentation
    D. Training for more epochs
69. What is L2 Regularization also known as?
    A. Lasso
    B. Dropout
    C. Ridge
    D. Early Stopping
70. What technique randomly deactivates a subset of neurons during each training iteration to prevent co-adaptation?
    A. L1 Regularization
    B. L2 Regularization
    C. Dropout
    D. Early Stopping
71. When does underfitting occur?
    A. When a model learns the training data too well.
    B. When a model is too simple to capture the underlying structure of the data.
    C. When a model generalizes perfectly to new data.
    D. When training time is too long.
72. How can underfitting be recognized?
    A. High training accuracy and low validation accuracy.
    B. Training loss continues to decrease while validation loss increases.
    C. Both training and validation accuracy remain low.
    D. The model performs extremely well on known data.
73. Which of the following is a strategy to fix underfitting?
    A. Reduce model complexity.
    B. Increase regularization.
    C. Train for a shorter duration.
    D. Increase model complexity (add more layers/neurons).
74. What fundamental principle in machine learning connects overfitting and underfitting?
    A. The Law of Large Numbers
    B. The Bias-Variance Tradeoff
    C. The Central Limit Theorem
    D. The No Free Lunch Theorem
75. What does "bias" refer to in the bias-variance tradeoff?
    A. Error introduced when a model learns the random noise.

B. Error introduced when a model oversimplifies the problem.
C. The model's ability to generalize.
D. The model's training speed.

76. What does "variance" refer to in the bias-variance tradeoff?
A. Error introduced when a model oversimplifies the problem.
B. Error introduced when a model learns the random noise in the training data.
C. The model's training accuracy.
D. The model's interpretability.

77. What is the ideal scenario in the bias-variance tradeoff?
A. High bias, high variance
B. Low bias, high variance
C. High bias, low variance
D. Low bias, low variance

78. What is a hyperparameter in deep learning?
A. A parameter learned during training.
B. A parameter whose value is set before the learning process begins.
C. A measure of model performance.
D. A type of activation function.

79. Which of the following is a key hyperparameter that significantly impacts model training?
A. Weights
B. Biases
C. Learning rate
D. Activations

80. What happens if the learning rate is too high?
A. The model will converge slowly.
B. The model may get stuck in a local minimum.
C. The model may "overshoot" the minimum and fail to converge.
D. The model will generalize perfectly.

81. What does "batch size" refer to?
A. The total number of training samples.
B. The number of training samples used in one iteration of model training.
C. The number of layers in the network.
D. The number of epochs.

82. What is a disadvantage of a small batch size?
A. Slower convergence.
B. Requires more memory.
C. May introduce more noise in the updates.
D. Smoother updates.

83. What is a disadvantage of a large batch size?
A. Faster convergence.
B. May lead to slower convergence.
C. Introduces more noise.
D. Helps escape local minima.

84. What do the number of layers and neurons per layer determine?
A. The learning rate
B. The batch size
C. The capacity of a neural network
D. The type of optimizer

85. What happens if a model has too few layers or neurons?

A. It will overfit the training data.
B. It will underfit the training data.
C. It will generalize perfectly.
D. It will train faster.

86. Which hyperparameter tuning method evaluates every possible combination of hyperparameter values in a defined grid?
A. Random Search
B. Bayesian Optimization
C. Grid Search
D. Genetic Algorithms

87. What is an advantage of Grid Search?
A. It is less computationally expensive.
B. It guarantees finding the best combination within the search space.
C. It covers a broader hyperparameter space for the same computational budget.
D. It is flexible.

88. What does Random Search evaluate?
A. Every possible combination of hyperparameters.
B. Random combinations of hyperparameters within specified ranges.
C. Only the most important hyperparameters.
D. Predefined combinations of hyperparameters.

89. What is an advantage of Random Search over Grid Search?
A. It guarantees finding the optimal combination.
B. It is more computationally expensive.
C. It may find good configurations with fewer evaluations.
D. It is less flexible.

90. What are the two major phases involved in building a deep neural network?
A. Data collection and data cleaning
B. Designing the network architecture and training the model
C. Model deployment and maintenance
D. Feature engineering and hyperparameter selection

91. What does designing the network architecture involve?
A. Calculating the loss function.
B. Planning the structure of the neural network (layers, neurons, etc.).
C. Updating the model's weights.
D. Monitoring performance during training.

92. What is the entry point of the neural network, taking in raw input data?
A. Hidden Layer
B. Output Layer
C. Input Layer
D. Convolutional Layer

93. What happens in the hidden layers of a neural network?
A. Final predictions are produced.
B. Raw input data is received.
C. The actual computation happens, applying weighted sums and activation functions.
D. Only linear transformations occur.

94. For multi-class classification, which activation function is commonly used in the output layer?
A. Sigmoid
B. ReLU

C. Tanh
D. Softmax

95. What is the MNIST dataset primarily used for?
A. Image generation
B. Classifying handwritten digits
C. Natural language processing
D. Time-series forecasting

96. How many pixels are in each MNIST image?
A. 32x32
B. 64x64
C. 28x28
D. 128x128

97. What is the purpose of monitoring performance during training a deep neural network?
A. To make the training process a black box.
B. To guide hyperparameter tuning and detect issues like overfitting.
C. To increase computational cost.
D. To randomly adjust weights.

98. Which two primary metrics are monitored throughout the training process?
A. Speed and memory
B. Loss and accuracy
C. Bias and variance
D. Input and output

99. What does a consistent decrease in both training and validation loss usually signal?
A. Overfitting
B. Underfitting
C. Healthy learning
D. Unstable learning

100. What technique involves halting training automatically when validation performance stops improving to prevent overfitting?
A. Data Augmentation
B. Regularization
C. Early Stopping
D. Model Simplification