1. **Supervised learning** is primarily used for:

    a) Finding patterns in unlabeled data
    b) Making predictions based on labeled data
    c) Creating new data instances
    d) Reducing dimensionality of data

2. **Unsupervised learning** is primarily used for:

    a) Making predictions based on labeled data
    b) Finding patterns in unlabeled data
    c) Creating new data instances
    d) Reducing dimensionality of data

3. **Reinforcement learning** differs from supervised and unsupervised learning primarily in:

    a) The type of data used
    b) The learning process, which involves an agent interacting with an environment
    c) The goal of the learning process
    d) The evaluation metric used

4. An **activation function** in a neural network is used to:

    a) Introduce non-linearity
    b) Increase the number of neurons
    c) Reduce the number of layers
    d) Normalize the input data

5. The **XOR problem** is significant because:

    a) It can be easily solved by a single perceptron
    b) It demonstrates the limitations of linear models
    c) It is a common problem in image recognition
    d) It is the basis for all deep learning algorithms

6. **Backpropagation** is a method used to:

    a) Initialize the weights of a neural network
    b) Calculate the output of a neural network
    c) Update the weights of a neural network based on the error
    d) Determine the optimal number of layers in a neural network

7. The **chain rule** is used in backpropagation to:

    a) Calculate the gradient of the loss function with respect to the weights
    b) Determine the optimal learning rate
    c) Prevent overfitting
    d) Initialize the weights of the neural network

8. A **batch** in gradient descent refers to:

    a) A group of neurons
    b) a group of training examples
    c) a single training example
    d) a hyperparameter

9. A **loss function** measures:

    a) The accuracy of a model
    b) the error between the predicted and actual values
    c) the computational efficiency of a model
    d) the number of parameters in a model

10. An **optimizer** is used to:

a) Initialize the weights of a neural network
b) Update the weights of a neural network to minimize the loss function
c) Determine the optimal number of layers in a neural network
d) Calculate the output of a neural network

11. **Underfitting** occurs when:

    a) A model is too complex for the data

    b) A model is too simple for the data

    c) The model has too many parameters

    d) The model has too few parameters

12. **Feature scaling** is important because:

    a) It improves the interpretability of the model

    b) It speeds up the training process

    c) It prevents overfitting

    d) It helps to normalize the data

13. A **fully connected layer** in a neural network means:

a) Every neuron in the layer is connected to every neuron in the previous layer

b) Only some neurons in the layer are connected to neurons in the previous layer

c) There are no connections between neurons in the layer

d) The layer has only one neuron

14. Which of the following activation functions is commonly used in the output layer of a neural network for binary classification problems?

    a) ReLU

    b) Sigmoid

    c) Tanh

    d) Softmax

15. The vanishing gradient problem is important to handle, because:

a) leads to slow convergence

b) leads to faster convergence

c) reaches zero

d) reaches one

16. Which of the following is NOT a common optimizer used in neural network training?

a) Gradient Descent

b) Adam

c) Momentum

d) Regularization

17. A learning rate that is too high can lead to:

a) Faster convergence

b) Oscillations and divergence

c) Underfitting

d) Overfitting

18. Which loss function is commonly used for regression problems?

    a) Cross-entropy

    b) Mean Squared Error (MSE)

    c) Hinge Loss

    d) Categorical Cross-entropy

19. Which of the following techniques can be used to prevent overfitting?

a) reduce the number of training examples

b) Early stopping

c) With more training iterations

d) Less training iterations

20. Which of the following techniques can be used to prevent overfitting?

a) reduce the number of training examples

b) Drop out

c) With more training iterations

d) Less training iterations

21. A model with high bias and low variance is likely to suffer from:

    a) Overfitting

    b) Underfitting

    c) Both overfitting and underfitting

    d) Neither overfitting nor underfitting

22. A model with Low bias and High variance is likely to suffer from:

    a) Overfitting

    b) Underfitting

    c) Both overfitting and underfitting

    d) Neither overfitting nor underfitting

23. Why is feature scaling important for neural networks?

a) To improve the convergence rate of the optimization algorithm

b) To prevent numerical instability

c) To improve the interpretability of the model

d) both a and b

24. In a fully connected layer, how many parameters are there (excluding biases)?

    a) Number of inputs * number of outputs

    b) Number of inputs + number of outputs

    c) Number of inputs * number of neurons in the layer

    d) Number of outputs * number of neurons in the layer

25. **A confusion matrix is used to:**

a) Evaluate regression models

b) Evaluate classification models

c) Visualize data distribution

d) Optimize hyperparameters

26. **Which of the following is NOT a component of a confusion matrix:**
    1. a) True Positive
    2. b) False Positive
    3. c) Mean Squared Error
    4. d) False Negative
27. **Precision is calculated as:**
    1. a) True Positives / (True Positives + True Negatives)
    2. b) True Positives / (True Positives + False Positives)
    3. c) True Positives / (True Positives + False Negatives)
    4. d) (True Positives + True Negatives) / Total Observations
28. **Recall is calculated as:**
    1. a) True Positives / (True Positives + True Negatives)
    2. b) True Positives / (True Positives + False Positives)
    3. c) True Positives / (True Positives + False Negatives)
    4. d) (True Positives + True Negatives) / Total Observations
29. **Which metric is more important when predicting a disease?**
    1. a) Precision
    2. b) Recall
    3. c) Accuracy
    4. Depends on the context
30. **The F1-score is the harmonic mean of:**
    1. a) Precision and Recall
    2. b) Accuracy and Precision
    3. c) Recall and Specificity
    4. d) Precision and Specificity
31. **A high false positive rate indicates:**
    1. a) Many actual positives are incorrectly classified as negatives
    2. b) Many actual negatives are incorrectly classified as positives
    3. c) The model is highly accurate
    4. d) The model is highly precise
32. **Which metric is more suitable for imbalanced datasets?**
    1. a) Accuracy
    2. b) Precision
    3. c) Recall
    4. d) F1-score