**1. What is regression analysis?**

Regression analysis is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It helps in predicting the dependent variable based on the independent variables.

---

**2. Explain the difference between linear and nonlinear regression.**

- **Linear Regression**: Assumes a straight-line relationship between the dependent and independent variables.
- **Nonlinear Regression**: Deals with situations where the relationship is not linear and can be modeled by a nonlinear equation.

---

**3. What is the difference between simple linear regression and multiple linear regression?**

- **Simple Linear Regression**: Models the relationship between a single independent variable and a dependent variable.
- **Multiple Linear Regression**: Models the relationship between two or more independent variables and a dependent variable.

---

**4. How is the performance of a regression model typically evaluated?**

Performance is usually evaluated using metrics like **Mean Squared Error (MSE)**, **R-squared**, and **Mean Absolute Error (MAE)**.

---

**5. What is overfitting in the context of regression models?**

Overfitting occurs when a model learns the noise in the training data instead of the underlying pattern, leading to poor performance on unseen data.

---

**6. What is logistic regression used for?**

Logistic regression is used for binary classification tasks, where the goal is to predict one of two possible outcomes.

---

**7. How does logistic regression differ from linear regression?**

- **Linear Regression**: Predicts a continuous value.
- **Logistic Regression**: Predicts a probability value between 0 and 1 using the sigmoid function.

---

**8. Explain the concept of odds ratio in logistic regression.**

The odds ratio in logistic regression indicates the odds of the dependent variable being 1 (success) relative to the odds of it being 0 (failure).

---

**9. What is the sigmoid function in logistic regression?**

The sigmoid function is used to map predicted values between 0 and 1, converting them into probabilities.

---

**10. How is the performance of a logistic regression model evaluated?**

Performance is evaluated using **Accuracy**, **Precision**, **Recall**, **F1 Score**, and **AUC-ROC curve**.

---

**11. What is a decision tree?**

A decision tree is a flowchart-like model used for classification and regression tasks. It splits data into subsets based on feature values to make predictions.

---

**12. How does a decision tree make predictions?**

A decision tree makes predictions by following the splits based on feature values until it reaches a leaf node, which provides the predicted value or class.

---

**13. What is entropy in the context of decision trees?**

Entropy is a measure of uncertainty or impurity in a dataset. A decision tree tries to reduce entropy at each split to create pure nodes.

---

**14. What is pruning in decision trees?**

Pruning is the process of removing unnecessary branches from the tree to prevent overfitting and improve generalization.

---

**15. How do decision trees handle missing values?**

Decision trees can handle missing values by using strategies like replacing them with the mean, median, or mode of the feature, or by making splits only on non-missing values.

---

**16. What is a support vector machine (SVM)?**

SVM is a supervised learning algorithm used for classification and regression. It finds the optimal hyperplane that separates different classes in the feature space.

---

**17. Explain the concept of margin in SVM.**

The margin in SVM is the distance between the separating hyperplane and the closest points from both classes, known as support vectors.

---

**18. What are support vectors in SVM?**

Support vectors are the data points that lie closest to the decision boundary. These points are crucial for defining the optimal hyperplane.

---

**19. How does SVM handle non-linearly separable data?**

SVM uses kernel functions to transform non-linearly separable data into a higher-dimensional space where it becomes linearly separable.

---

**20. What are the advantages of SVM over other classification algorithms?**

SVM is effective in high-dimensional spaces, is robust to overfitting, and works well for both linear and non-linear classification tasks.

---

**21. What is the Naïve Bayes algorithm?**

Naïve Bayes is a probabilistic classifier based on Bayes' theorem with strong (naïve) independence assumptions between features.

---

**22. Why is it called "Naïve" Bayes?**

It is called "Naïve" because it assumes that all features are independent, which is often not the case in real-world data.

---

**23. How does Naïve Bayes handle continuous and categorical features?**

- **Categorical features**: Naïve Bayes uses probability distributions (e.g., multinomial or Bernoulli) for categorical data.
- **Continuous features**: It assumes continuous features follow a normal distribution and uses Gaussian Naïve Bayes.

---

**24. Explain the concept of prior and posterior probabilities in Naïve Bayes.**

- **Prior probability**: The probability of a class before observing any data.
- **Posterior probability**: The probability of a class given the observed data.

---

**25. What is Laplace smoothing and why is it used in Naïve Bayes?**

Laplace smoothing is a technique to handle zero probabilities by adding a small constant to each count, ensuring no probability is zero.

---

**26. Can Naïve Bayes be used for regression tasks?**

Naïve Bayes is typically used for classification tasks, but it can be adapted for regression with some modifications.

---

**27. How do you handle missing values in Naïve Bayes?**

Naïve Bayes can handle missing values by ignoring the missing features during the probability computation or by using the mean/mode for imputation.

---

**28. What are some common applications of Naïve Bayes?**

Naïve Bayes is commonly used in spam detection, sentiment analysis, and document classification tasks.

---

**29. Explain the concept of feature independence assumption in Naïve Bayes.**

Naïve Bayes assumes that all features are independent of each other, which simplifies the computation of joint probabilities.

---

**30. How does Naïve Bayes handle categorical features with a large number of categories?**

Naïve Bayes handles categorical features with many categories by calculating the probability for each category and using Laplace smoothing if needed.

---

**31. What is the curse of dimensionality, and how does it affect machine learning algorithms?**

The curse of dimensionality refers to the challenges and inefficiencies that arise when the number of features increases, such as increased computational cost and overfitting.

---

**32. Explain the bias-variance tradeoff and its implications for machine learning models.**

The bias-variance tradeoff involves balancing underfitting (high bias) and overfitting (high variance). A model with high bias is too simple, while high variance means it is too complex.

---

**33. What is cross-validation, and why is it used?**

Cross-validation is a technique used to assess the performance of a model by training and testing it on different subsets of the data to ensure robustness.

---

**34. Explain the difference between parametric and non-parametric machine learning algorithms.**

- **Parametric**: Make assumptions about the underlying data distribution (e.g., linear regression).
- **Non-parametric**: Do not make assumptions about data distribution (e.g., k-NN, decision trees).

---

**35. What is feature scaling, and why is it important in machine learning?**

Feature scaling is the process of normalizing or standardizing features so that they are on a similar scale. It is important for algorithms like SVM, k-NN, and gradient descent.

---

**36. What is regularization, and why is it used in machine learning?**

Regularization is a technique used to prevent overfitting by adding a penalty term to the loss function (e.g., L1 or L2 regularization).

---

**37. Explain the concept of ensemble learning and give an example.**

Ensemble learning combines multiple models to improve prediction accuracy. Examples include **Random Forests** and **Gradient Boosting**.

---

**38. What is the difference between bagging and boosting?**

- **Bagging**: Combines multiple weak models (e.g., Random Forest) to reduce variance.
- **Boosting**: Sequentially combines models, with each model focusing on the errors of the previous ones (e.g., AdaBoost, XGBoost).

---

**39. What is the difference between a generative model and a discriminative model?**

- **Generative models**: Model the distribution of each class (e.g., Naïve Bayes).
- **Discriminative models**: Model the decision boundary between classes (e.g., SVM, Logistic Regression).

---

**40. Explain the concept of batch gradient descent and stochastic gradient descent.**

- **Batch Gradient Descent**: Computes the gradient using the entire dataset.

- **Stochastic Gradient Descent**: Computes the gradient using a single data point at each iteration.

---

**41. What is the K-nearest neighbors (KNN) algorithm, and how does it work?**

KNN is a classification algorithm that predicts the class of a data point based on the majority class of its k-nearest neighbors.

---

**42. What are the disadvantages of the K-nearest neighbors algorithm?**

KNN is computationally expensive, especially with large datasets, and it struggles with high-dimensional data.

---

**43. Explain the concept of one-hot encoding and its use in machine learning.**

One-hot encoding converts categorical variables into binary vectors, where each category is represented by a unique binary code.

---

**44. What is feature selection, and why is it important in machine learning?**

Feature selection involves selecting the most relevant features to improve model performance and reduce complexity.

---

**45. Explain the concept of cross-entropy loss and its use in classification tasks.**

Cross-entropy loss measures the difference between the predicted probability distribution and the true distribution, commonly used in classification tasks.

---

**46. What is the difference between batch learning and online learning?**

- **Batch Learning**: The model is trained on the entire dataset at once.
- **Online Learning**: The model is trained incrementally, updating with each new data point.

---

**47. Explain the concept of grid search and its use in hyperparameter tuning.**

Grid search is an exhaustive search method for finding the optimal hyperparameters by testing all combinations of a predefined set of hyperparameters.

---

**48. What are the advantages and disadvantages of decision trees?**

- **Advantages**: Easy to understand, interpretable, and can handle both numerical and categorical data.

- **Disadvantages**: Prone to overfitting and unstable to small variations in data.

---

**49. What is the difference between L1 and L2 regularization?**

- **L1 Regularization (Lasso)**: Adds the absolute values of coefficients to the loss function, encouraging sparsity.
- **L2 Regularization (Ridge)**: Adds the squared values of coefficients, preventing large coefficients.

---

**50. What are some common preprocessing techniques used in machine learning?**

Common techniques include **scaling**, **normalization**, **missing value imputation**, **encoding categorical variables**, and **feature extraction**.