Regression Analysis

1. **What is regression analysis?**
   Regression analysis is a statistical method used to examine the relationship between a dependent variable and one or more independent variables. It helps in predicting the value of the dependent variable based on the values of the independent variables.
2. **Explain the difference between linear and nonlinear regression:**
   - **Linear regression** assumes a linear relationship between the dependent and independent variables.
   - **Nonlinear regression** assumes a more complex, nonlinear relationship, often modeled using curves or higher-order terms.
3. **What is the difference between simple linear regression and multiple linear regression?**
   - **Simple linear regression** involves one independent variable and one dependent variable.
   - **Multiple linear regression** involves two or more independent variables and one dependent variable.
4. **How is the performance of a regression model typically evaluated?**
   Common metrics include **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)**, **Mean Absolute Error (MAE)**, and **R-squared (R^2)**.
5. **What is overfitting in the context of regression models?**
   Overfitting occurs when a model learns the training data too well, capturing noise and outliers, which reduces its performance on unseen data.

Logistic Regression

6. **What is logistic regression used for?**
   Logistic regression is used for binary classification tasks, predicting the probability of an event occurring (e.g., yes/no, true/false).
7. **How does logistic regression differ from linear regression?**
   Logistic regression predicts probabilities using a logistic function, while linear regression predicts continuous values.
8. **Explain the concept of odds ratio in logistic regression:**
   The odds ratio measures the association between an independent variable and the outcome. It represents the change in odds of the outcome for a one-unit change in the predictor.

9. **What is the sigmoid function in logistic regression?**
   The sigmoid function is an S-shaped curve that maps any real-valued number into a value between 0 and 1, representing probabilities.
10. **How is the performance of a logistic regression model evaluated?**
   Common metrics include **Accuracy**, **Precision**, **Recall**, **F1-Score**, and **ROC-AUC**.

Decision Trees

11. **What is a decision tree?**
   A decision tree is a tree-like model that splits data into branches based on feature values to make predictions.
12. **How does a decision tree make predictions?**
   It follows a path from the root to a leaf node, making decisions based on feature values at each split.
13. **What is entropy in the context of decision trees?**
   Entropy measures the impurity or randomness in a dataset. Decision trees aim to minimize entropy to create pure splits.
14. **What is pruning in decision trees?**
   Pruning is the process of removing unnecessary branches to reduce overfitting and improve generalization.
15. **How do decision trees handle missing values?**
   They can handle missing values by using surrogate splits or imputing values based on the most common class or mean.

Support Vector Machines (SVM)

16. **What is a support vector machine (SVM)?**
   SVM is a supervised learning algorithm used for classification and regression tasks by finding the optimal hyperplane that separates classes.
17. **Explain the concept of margin in SVM:**
   The margin is the distance between the hyperplane and the nearest data points (support vectors). SVM aims to maximize this margin.
18. **What are support vectors in SVM?**
   Support vectors are the data points closest to the hyperplane that influence its position and orientation.
19. **How does SVM handle non-linearly separable data?**
   It uses kernel functions (e.g., polynomial, radial basis function) to transform data into a higher-dimensional space where it becomes separable.

20. **What are the advantages of SVM over other classification algorithms?**
    SVM is effective in high-dimensional spaces, robust to overfitting, and works well with clear margin separation.

Naïve Bayes

21. **What is the Naïve Bayes algorithm?**
    Naïve Bayes is a probabilistic classifier based on Bayes' theorem, assuming independence between features.
22. **Why is it called "Naïve" Bayes?**
    It is called "naïve" because it assumes that all features are independent of each other, which is often not true in real-world data.
23. **How does Naïve Bayes handle continuous and categorical features?**
    - For categorical features, it uses frequency counts.
    - For continuous features, it assumes a probability distribution (e.g., Gaussian).
24. **Explain the concept of prior and posterior probabilities in Naïve Bayes:**
    - **Prior probability** is the initial probability of an event before observing data.
    - **Posterior probability** is the updated probability after considering the evidence.
25. **What is Laplace smoothing and why is it used in Naïve Bayes?**
    Laplace smoothing adds a small constant to avoid zero probabilities when a feature value does not appear in the training data.
26. **Can Naïve Bayes be used for regression tasks?**
    No, Naïve Bayes is primarily used for classification tasks.
27. **How do you handle missing values in Naïve Bayes?**
    Missing values are ignored during probability calculations, or imputed using mean/mode.
28. **What are some common applications of Naïve Bayes?**
    Spam filtering, sentiment analysis, and document classification.
29. **Explain the concept of feature independence assumption in Naïve Bayes.**
    It assumes that all features contribute independently to the probability of the outcome, which simplifies calculations but may not hold in reality.
30. **How does Naïve Bayes handle categorical features with a large number of categories?**
    It may struggle due to sparse data, but Laplace smoothing can help mitigate this issue.

General Machine Learning

31. **What is the curse of dimensionality, and how does it affect machine learning algorithms?**
    As the number of features increases, data becomes sparse, making it harder for algorithms to find patterns.
32. **Explain the bias-variance tradeoff and its implications for machine learning models:**
    - **Bias** is the error due to overly simplistic assumptions.
    - **Variance** is the error due to sensitivity to small fluctuations in the training set.
      A good model balances both.
33. **What is cross-validation, and why is it used?**
    Cross-validation is a technique to evaluate model performance by partitioning data into subsets, training on some, and validating on others.
34. **Explain the difference between parametric and non-parametric machine learning algorithms:**
    - **Parametric** algorithms assume a fixed number of parameters (e.g., linear regression).
    - **Non-parametric** algorithms do not make strong assumptions about the data (e.g., decision trees).
35. **What is feature scaling, and why is it important in machine learning?**
    Feature scaling normalizes data to a standard range, improving the performance of algorithms sensitive to feature magnitudes (e.g., SVM, KNN).
36. **What is regularization, and why is it used in machine learning?**
    Regularization adds a penalty to the loss function to prevent overfitting by discouraging large coefficients.
37. **Explain the concept of ensemble learning and give an example:**
    Ensemble learning combines multiple models to improve performance (e.g., Random Forest, which combines decision trees).
38. **What is the difference between bagging and boosting?**
    - **Bagging** trains models independently and averages their predictions.
    - **Boosting** trains models sequentially, focusing on errors from previous models.
39. **What is the difference between a generative model and a discriminative model?**
    - **Generative models** learn the joint probability distribution of inputs and outputs (e.g., Naïve Bayes).
    - **Discriminative models** learn the conditional probability of outputs given inputs (e.g., logistic regression).

40. **Explain the concept of batch gradient descent and stochastic gradient descent:**
    - **Batch gradient descent** updates model parameters using the entire dataset.
    - **Stochastic gradient descent** updates parameters using one data point at a time, making it faster but noisier.
41. **What is the K-nearest neighbors (KNN) algorithm, and how does it work?**
    KNN predicts the class or value of a data point based on the majority class or average of its k-nearest neighbors.
42. **What are the disadvantages of the K-nearest neighbors algorithm?**
    It is computationally expensive, sensitive to irrelevant features, and requires careful choice of k.
43. **Explain the concept of one-hot encoding and its use in machine learning:**
    One-hot encoding converts categorical variables into binary vectors, making them suitable for machine learning algorithms.
44. **What is feature selection, and why is it important in machine learning?**
    Feature selection identifies the most relevant features, reducing dimensionality and improving model performance.
45. **Explain the concept of cross-entropy loss and its use in classification tasks:**
    Cross-entropy loss measures the difference between predicted and actual probabilities, guiding the model to improve predictions.
46. **What is the difference between batch learning and online learning?**
    - **Batch learning** trains the model on the entire dataset at once.
    - **Online learning** updates the model incrementally as new data arrives.
47. **Explain the concept of grid search and its use in hyperparameter tuning:**
    Grid search systematically tests hyperparameter combinations to find the best-performing model.
48. **What are the advantages and disadvantages of decision trees?**
    - **Advantages:** Easy to interpret, handle both numerical and categorical data.
    - **Disadvantages:** Prone to overfitting, sensitive to small data changes.
49. **What is the difference between L1 and L2 regularization?**
    - **L1 regularization** adds the absolute value of coefficients to the loss function, encouraging sparsity.
    - **L2 regularization** adds the squared value of coefficients, discouraging large values.
50. **What are some common preprocessing techniques used in machine learning?**
    Normalization, standardization, handling missing values, and encoding categorical variables.

Advanced Topics

51. **What is the difference between a parametric and non-parametric algorithm? Give examples of each:**
    - **Parametric:** Linear regression, logistic regression.
    - **Non-parametric:** Decision trees, KNN.
52. **Explain the bias-variance tradeoff and how it relates to model complexity:**
    As model complexity increases, bias decreases but variance increases, leading to overfitting.
53. **What are the advantages and disadvantages of using ensemble methods like random forests?**
    - **Advantages:** High accuracy, handles overfitting.
    - **Disadvantages:** Computationally expensive, less interpretable.
54. **Explain the difference between bagging and boosting:**
    Bagging reduces variance, while boosting reduces bias.
55. **What is the purpose of hyperparameter tuning in machine learning?**
    To optimize model performance by finding the best hyperparameter values.
56. **What is the difference between regularization and feature selection?**
    Regularization penalizes model complexity, while feature selection removes irrelevant features.
57. **How does the Lasso (L1) regularization differ from Ridge (L2) regularization?**
    Lasso encourages sparsity by shrinking some coefficients to zero, while Ridge shrinks all coefficients proportionally.
58. **Explain the concept of cross-validation and why it is used:**
    Cross-validation assesses model performance by partitioning data into training and validation sets multiple times.
59. **What are some common evaluation metrics used for regression tasks?**
    MSE, RMSE, MAE, and R-squared.
60. **How does the K-nearest neighbors (KNN) algorithm make predictions?**
    It finds the k-nearest data points and predicts based on their majority class or average value.
61. **What is the curse of dimensionality, and how does it affect machine learning algorithms?**
    High-dimensional data becomes sparse, making it harder for algorithms to find meaningful patterns.
62. **What is feature scaling, and why is it important in machine learning?**
    It normalizes data to a standard range, improving algorithm performance.
63. **How does the Naïve Bayes algorithm handle categorical features?**
    It uses frequency counts to estimate probabilities.

64. **Explain the concept of prior and posterior probabilities in Naïve Bayes:**
    Prior is the initial probability, and posterior is the updated probability after observing data.
65. **What is Laplace smoothing, and why is it used in Naïve Bayes?**
    It prevents zero probabilities by adding a small constant to frequency counts.
66. **Can Naïve Bayes handle continuous features?**
    Yes, by assuming a probability distribution (e.g., Gaussian).
67. **What are the assumptions of the Naïve Bayes algorithm?**
    Feature independence and equal importance of features.
68. **How does Naïve Bayes handle missing values?**
    It ignores missing values during probability calculations.
69. **What are some common applications of Naïve Bayes?**
    Spam filtering, sentiment analysis, and document classification.
70. **Explain the difference between generative and discriminative models:**
    Generative models learn the joint distribution, while discriminative models learn the conditional distribution.
71. **How does the decision boundary of a Naïve Bayes classifier look like for binary classification tasks?**
    It is often linear but can be nonlinear depending on the data distribution.
72. **What is the difference between multinomial Naïve Bayes and Gaussian Naïve Bayes?**
    Multinomial Naïve Bayes is used for discrete data, while Gaussian Naïve Bayes is used for continuous data.
73. **How does Naïve Bayes handle numerical instability issues?**
    By using log probabilities to avoid underflow.
74. **What is the Laplacian correction, and when is it used in Naïve Bayes?**
    It is another term for Laplace smoothing, used to handle zero probabilities.
75. **Can Naïve Bayes be used for regression tasks?**
    No, it is designed for classification.
76. **Explain the concept of conditional independence assumption in Naïve Bayes:**
    It assumes that features are independent given the class label.
77. **How does Naïve Bayes handle categorical features with a large number of categories?**
    It may struggle due to sparse data, but Laplace smoothing helps.
78. **What are some drawbacks of the Naïve Bayes algorithm?**
    It assumes feature independence, which may not hold in real-world data.
79. **Explain the concept of smoothing in Naïve Bayes:**
    Smoothing prevents zero probabilities by adding a small constant to frequency counts.

80. **How does Naïve Bayes handle imbalanced datasets?**
It can handle imbalanced datasets by adjusting class priors or using techniques like SMOTE.