ML Assignment -2

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.

2. Explain the difference between linear and nonlinear regression:
Linear regression assumes a straight-line relationship, while nonlinear regression models more complex, curved relationships.

3. What is the difference between simple linear regression and multiple linear regression?
Simple linear regression uses one independent variable, while multiple linear regression uses more than one.

4. How is the performance of a regression model typically evaluated?
It is evaluated using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.

5. What is overfitting in the context of regression models?
Overfitting occurs when a model learns the noise in the training data, leading to poor performance on unseen data.

6. What is logistic regression used for?
Logistic regression is used for binary classification tasks, predicting probabilities of a binary outcome.

7. How does logistic regression differ from linear regression?
Logistic regression outputs probabilities using the sigmoid function, while linear regression predicts continuous values.

8. Explain the concept of odds ratio in logistic regression:
The odds ratio is the ratio of the odds of an event occurring vs. not occurring, used to interpret logistic regression coefficients.

9. What is the sigmoid function in logistic regression?
The sigmoid function maps any real-valued number to a value between 0 and 1, representing a probability.

10. How is the performance of a logistic regression model evaluated?
It is typically evaluated using metrics like accuracy, precision, recall, F1 score, and AUC-ROC.

11. What is a decision tree?
A decision tree is a model that makes decisions based on feature splits, forming a tree-like structure of choices.

12. How does a decision tree make predictions?
It splits data based on feature values and assigns labels based on the majority class at the leaves.

13. What is entropy in the context of decision trees?
Entropy is a measure of uncertainty or impurity used to decide where to split the data.

14. What is pruning in decision trees?
Pruning involves removing 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 surrogate splits or assigning a default class.

16. What is a support vector machine (SVM)?
SVM is a classification algorithm that separates classes using a hyperplane that maximizes the margin between them.

17. Explain the concept of margin in SVM:
The margin is the distance between the support vectors and the decision boundary, with a larger margin improving generalization.

18. What are support vectors in SVM?
Support vectors are the data points closest to the decision boundary, critical in defining the hyperplane.

19. How does SVM handle non-linearly separable data?
SVM uses kernel functions to map data into higher-dimensional spaces for better separation.

20. What are the advantages of SVM over other classification algorithms?
SVM is effective in high-dimensional spaces and can handle non-linear decision boundaries using kernels.

21. What is the Naïve Bayes algorithm?
Naïve Bayes is a probabilistic classifier based on Bayes' theorem, assuming feature independence.

22. Why is it called "Naïve" Bayes?
It is called "naïve" because it assumes that all features are independent, which is rarely true in real-world data.

23. How does Naïve Bayes handle continuous and categorical features?
Naïve Bayes uses Gaussian distribution for continuous features and categorical probability distributions for categorical features.

24. Explain the concept of prior and posterior probabilities in Naïve Bayes:
Prior probability is the initial belief about the data, while posterior probability is the updated belief after seeing the data.

25. What is Laplace smoothing and why is it used in Naïve Bayes?
Laplace smoothing adds a small constant to probabilities to avoid zero probabilities for unseen events.

26. Can Naïve Bayes be used for regression tasks?
Naïve Bayes is generally used for classification tasks, but it can be adapted for regression in some cases.

27. How do you handle missing values in Naïve Bayes?
Missing values can be handled by ignoring the feature or assigning a default value.

28. What are some common applications of Naïve Bayes?
Applications include spam filtering, text classification, and sentiment analysis.

29. Explain the concept of feature independence assumption in Naïve Bayes:
It assumes that all features are independent, simplifying the model but reducing its accuracy with correlated features.

30. How does Naïve Bayes handle categorical features with a large number of categories?
Naïve Bayes can handle large categories but may suffer from high variance or sparsity, requiring techniques like smoothing.

31. What is the curse of dimensionality, and how does it affect machine learning algorithms?
The curse of dimensionality refers to the difficulty in managing data as the number of features increases, leading to sparse data.

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

The bias-variance tradeoff balances between underfitting (high bias) and overfitting (high variance), influencing model complexity.

33. What is cross-validation, and why is it used?
Cross-validation is a technique for assessing model performance by splitting data into multiple training and testing sets.

34. Explain the difference between parametric and non-parametric machine learning algorithms:
Parametric algorithms assume a fixed form for the model, while non-parametric algorithms make fewer assumptions about data structure.

35. What is feature scaling, and why is it important in machine learning?
Feature scaling normalizes features to a similar range, improving the performance of algorithms that are sensitive to feature magnitudes.

36. What is regularization, and why is it used in machine learning?
Regularization penalizes large model parameters to prevent overfitting and improve generalization.

37. Explain the concept of ensemble learning and give an example:
Ensemble learning combines multiple models to improve accuracy, e.g., Random Forests.

38. What is the difference between bagging and boosting?
Bagging reduces variance by training multiple models independently, while boosting reduces bias by iteratively improving weak models.

39. What is the difference between a generative model and a discriminative model?
Generative models model the data distribution, while discriminative models directly model the decision boundary.

40. Explain the concept of batch gradient descent and stochastic gradient descent:
Batch gradient descent computes gradients over the entire dataset, while stochastic gradient descent uses one data point at a time.

41. What is the K-nearest neighbors (KNN) algorithm, and how does it work?
KNN is a non-parametric algorithm that classifies data based on the majority label of its K closest neighbors.

42. What are the disadvantages of the K-nearest neighbors algorithm?
Disadvantages include high computation cost, sensitivity to irrelevant features, and the need for feature scaling.

43. Explain the concept of one-hot encoding and its use in machine learning:
One-hot encoding converts categorical features into binary vectors, enabling machine learning algorithms to handle categorical data.

44. What is feature selection, and why is it important in machine learning?
Feature selection identifies and removes irrelevant or redundant features to improve model performance and reduce overfitting.

45. Explain the concept of cross-entropy loss and its use in classification tasks:
Cross-entropy loss measures the difference between predicted probabilities and actual class labels, commonly used in classification tasks.

46. What is the difference between batch learning and online learning?
Batch learning trains the model on the entire dataset at once, while online learning updates the model incrementally with each data point.

47. Explain the concept of grid search and its use in hyperparameter tuning:
Grid search exhaustively searches through predefined hyperparameter combinations to find the best model configuration.

48. What are the advantages and disadvantages of decision trees?
Advantages: easy to interpret, handle both numerical and categorical data.
Disadvantages: prone to overfitting and unstable with small changes in data.

49. What is the difference between L1 and L2 regularization?
L1 regularization (Lasso) promotes sparsity by setting some coefficients to zero, while L2 regularization (Ridge) penalizes large coefficients without eliminating them.

50. What are some common preprocessing techniques used in machine learning?
Common techniques include normalization, standardization, missing value imputation, and encoding categorical variables.

51. What is the difference between a parametric and non-parametric algorithm? Give examples of each:
Parametric algorithms, like Linear Regression, assume a fixed model form. Non-parametric algorithms, like KNN, make fewer assumptions.

52. Explain the bias-variance tradeoff and how it relates to model complexity:
Increasing model complexity can reduce bias but increase variance, while simpler models reduce variance but increase bias.

53. What are the advantages and disadvantages of using ensemble methods like random forests?
Advantages: improved accuracy and robustness. Disadvantages: computationally expensive and harder to interpret.

54. What is the difference between bagging and boosting?
Bagging trains models independently on random subsets of the data, while boosting trains models sequentially, focusing on errors made by previous models.

55. What is the purpose of hyperparameter tuning in machine learning?
Hyperparameter tuning optimizes the model's performance by selecting the best hyperparameters through methods like grid search or random search.

56. What is the difference between regularization and feature selection?
Regularization penalizes large coefficients to reduce overfitting, while feature selection removes irrelevant features to improve model efficiency.

57. How does the Lasso (L1) regularization differ from Ridge (L2) regularization?
Lasso (L1) promotes sparsity by shrinking some coefficients to zero, while Ridge (L2) penalizes large coefficients but does not set them to zero.

58. What is a confusion matrix, and how is it used in classification tasks?
A confusion matrix is a table that summarizes the performance of a classification algorithm by showing the true positive, false positive, true negative, and false negative results.

59. Explain the concept of recall in classification problems:
Recall is the ratio of correctly predicted positive observations to all actual positives. It measures the ability of a model to capture all relevant cases.

60. What is precision in classification tasks?
Precision is the ratio of correctly predicted positive observations to all predicted positives. It measures the accuracy of the positive predictions.

61. What is the F1-score, and why is it important?
The F1-score is the harmonic mean of precision and recall, providing a balanced measure of performance, especially when classes are imbalanced.

62. What is the ROC curve, and how is it used in evaluating classification models?
The ROC (Receiver Operating Characteristic) curve plots the true positive rate against the false positive rate, helping evaluate the trade-off between sensitivity and specificity.

63. What is AUC-ROC?
AUC-ROC (Area Under the Curve - Receiver Operating Characteristic) is a metric that summarizes the performance of a classification model across all classification thresholds.

64. What is the concept of multicollinearity, and how does it affect regression models?
Multicollinearity occurs when independent variables in a regression model are highly correlated, leading to unstable estimates and inflated standard errors.

65. What is regularization in linear regression?
Regularization in linear regression (Lasso, Ridge) adds a penalty to the loss function to avoid overfitting by shrinking model coefficients.

66. What are hyperparameters in machine learning?
Hyperparameters are parameters set before training the model, such as learning rate, number of trees in a forest, or regularization strength.

67. How does the learning rate affect training in gradient descent?
The learning rate controls the step size of the updates during training; if it's too large, the model might overshoot the optimal solution, while a small value may result in slow convergence.

68. What is a learning curve in machine learning?
A learning curve shows how a model's performance improves with more training data or iterations, helping to assess whether the model is overfitting or underfitting.

69. Explain the difference between supervised and unsupervised learning:
Supervised learning involves training a model with labeled data, while unsupervised learning uses data without labels to find patterns or groupings.

70. What is clustering, and what are its common algorithms?
Clustering is the process of grouping similar data points together. Common algorithms include K-Means, DBSCAN, and hierarchical clustering.

71. What is dimensionality reduction, and why is it used?
Dimensionality reduction reduces the number of input features while retaining the important information, improving computational efficiency and model performance.

72. What is PCA (Principal Component Analysis), and how does it work?
PCA is a technique for dimensionality reduction that transforms data into a set of orthogonal components that capture the most variance in the data.

73. What is t-SNE (t-Distributed Stochastic Neighbor Embedding)?
t-SNE is a technique for dimensionality reduction that focuses on preserving the local structure of the data, commonly used for visualization.

74. What is the difference between classification and regression?
Classification involves predicting discrete labels, while regression predicts continuous numerical values.

75. What are some common applications of machine learning?
Common applications include recommendation systems, fraud detection, image recognition, natural language processing, and autonomous driving.

76. What is reinforcement learning?
Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with the environment and receiving feedback in the form of rewards or penalties.

77. What is the difference between model training and model evaluation?
Training is the process of learning from the data to fit the model, while evaluation assesses the model's performance on unseen data using metrics like accuracy or F1-score.

78. What is an outlier, and how does it affect machine learning models?
An outlier is an observation that significantly differs from the rest of the data. Outliers can distort model training, leading to inaccurate predictions.

79. What is the curse of dimensionality in clustering?
The curse of dimensionality refers to the difficulty of clustering high-dimensional data, where distance measures become less meaningful, and computational cost increases.

80. What are some techniques to handle imbalanced datasets in classification tasks?
Techniques include oversampling the minority class (e.g., SMOTE),

undersampling the majority class, or using algorithmic approaches like class weighting or cost-sensitive learning.