**Section – A (20 questions )**

**1. What does R-squared measure in the context of regression models?**

a) The proportion of variance in the dependent variable explained by the independent variables.

b) The difference between predicted and actual values.

c) The accuracy of the classification model.

d) The degree of multicollinearity among independent variables.

**2. Which of the following metrics is used to assess the precision of a classification model?**

a) R-squared b) Confusion Matrix c) Mean Squared Error d) Accuracy

**3. What is the purpose of using logistic regression in machine learning?**

a) To predict continuous outcomes. b) To perform dimensionality reduction.

c) To classify data into discrete categories. d) To handle missing values in the dataset.

**4. Which evaluation metric assesses the proportion of correctly classified instances out of all instances in a classification model?**

a) R-squared b) Mean Squared Error c) Precision d) Root Mean Squared Error

**5. What is the primary purpose of using the activation function in neural networks?**

a) To transform the input features into a higher-dimensional space. b) To calculate the cost function.

c) To introduce non-linearity into the model. d) To regularize the model parameters.

**6. Which of the following is NOT a type of kernel function commonly used in Support Vector Machines?**

a) Gaussian kernel b) Polynomial kernel c) Exponential kernel d) Sigmoid kernel

**7. What does the term "one-vs-all" mean in the context of multiclass classification with logistic regression?**

a) Training multiple models for each class. b) Combining all classes into one category.

c) Using one classifier for all classes simultaneously. d) Handling binary classification problems only.

**8. Which of the following evaluation metrics measures the model's ability to recall all positive instances correctly?**

a) Precision b) Recall c) F1-score d) Specificity

**9. In linear regression, what does the residual represent?**

a) The difference between the predicted and actual values. b) The slope of the regression line.

c) The intercept of the regression line. d) The variance explained by the model.

**10. What is the main drawback of using decision trees for classification tasks?**

a) Tendency to underfit the data. b) Prone to multicollinearity.

c) Susceptible to overfitting. d) Limited to linear decision boundaries.

**11. Which metric is used to evaluate the performance of a binary classification model when the cost of false positives is high?**

a) Precision b) Recall c) F1-score d) Accuracy

**12. What is the primary purpose of using cross-validation in machine learning?**

a) To maximize the training accuracy of the model.

b) To minimize the computational cost of model training.

c) To assess the generalization performance of the model.

d) To optimize the hyperparameters of the model.

**13. How does multicollinearity affect the performance of linear regression models?**

a) It improves the model's ability to capture non-linear relationships.

b) It reduces the stability of the coefficient estimates and may lead to inflated standard errors.

c) It increases the model's interpretability by simplifying the relationship between predictors and the target variable.

d) It has no impact on the performance of linear regression models.

**14. What is the significance of the intercept term in linear regression models?**

a) It represents the slope of the regression line.

b) It determines the direction of the regression line.

c) It adjusts the magnitude of the coefficients.

d) It accounts for the baseline value of the dependent variable.

**15. Explain the concept of precision in the context of classification models.**

a) It measures the proportion of true positive predictions out of all positive instances.

b) It quantifies the ability of the model to correctly identify negative instances.

c) It assesses the model's performance in predicting true negative instances.

d) It evaluates the balance between true positive and false positive predictions.

**16. What does the term "sigmoid function" refer to in logistic regression?**

a) A function used to calculate the cost of misclassifications.

b) A function used to normalize the input features.

c) A function used to introduce non-linearity into the model.

d) A function used to transform the linear combination of features into probabilities.

**17. Which of the following is NOT a characteristic of decision trees?**

a) They are robust to outliers. b) They can capture non-linear relationships.

c) They are sensitive to feature scaling. d) They can handle both numerical and categorical data.

**18. What is the primary disadvantage of using the mean squared error (MSE) as an evaluation metric for regression models?**

a) It is not sensitive to outliers. b) It cannot handle non-linear relationships.

c) It may be heavily influenced by extreme values.

d) It does not provide a clear interpretation of model performance.

**19. Which of the following evaluation metrics is used to assess the goodness of fit of a regression model?**

a) Confusion Matrix b) Precision c) R-squared d) F1-score

**20. In logistic regression, what is the purpose of the activation function?**

a) To calculate the cost function.

b) To transform the linear combination of features into probabilities.

c) To regularize the model. d) To normalize the input features.

**Section – B (15 questions – 2 marks each)**

**21. Which of the following is NOT a common kernel function used in Support Vector Machines?**

a) Linear kernel b) Polynomial kernel c) Sigmoid kernel d) Exponential kernel

**22. What is the primary goal of using regularization techniques in machine learning models?**

a) To increase the model complexity b) To reduce the risk of overfitting

c) To decrease the computational cost d) To improve model interpretability

**23. Which metric is used to evaluate the performance of a binary classification model when the cost of false negatives is high?**

a) Precision b) Recall c) F1-score d) Specificity

**24. In logistic regression, what does the odds ratio represent?**

a) The ratio of the probability of the event occurring to the probability of it not occurring

b) The difference between the predicted and actual values

c) The slope of the regression line d) The probability of the event occurring

**25. Discuss the significance of feature scaling in machine learning models?**

a) Feature scaling ensures that all features contribute equally to the model's learning process and improves convergence speed and model performance.

b) Feature scaling reduces the dimensionality of the dataset, making it easier to visualize and interpret the results.

c) Feature scaling helps in reducing the computational complexity of the model, leading to faster training times.

d) Feature scaling is not necessary for machine learning models and can sometimes introduce unnecessary complexity.

**Answer: Common techniques include Min-Max scaling, Standardization (Z-score scaling), and Robust scaling.**  
  
**26. What is the primary objective of k-means clustering?**

a) To classify data points into predefined categories b) To group similar data points together

c) To predict the value of a continuous variable d) To reduce the dimensionality of the data

**27. What is the primary objective of hyperparameter tuning in Machine Learning?**

a) To improve model performance b) To preprocess the data

c) To visualize the data d) To understand the data

**28. What is the purpose of the continue statement in Python?**

a) To exit the loop completely

b) To skip the current iteration of the loop and continue with the next iteration

c) To skip the next iteration of the loop

d) To resume the execution of the loop from the beginning

**29. What is the main advantage of using gradient boosting algorithms?**

a) They are resistant to overfitting b) They can handle missing data

c) They combine weak learners to create a strong learner d) They are computationally efficient

**29. What is the primary objective of cross-validation in Machine Learning?**

a) To train the model on multiple subsets of data b) To test the model on multiple subsets of data

c) To optimize hyperparameters d) All of the above

**30. Which of the following is a dimensionality reduction technique?**

a) K-means clustering b) Decision trees

c) Principal Component Analysis (PCA) d) Gradient Boosting

**31. Which of the following is NOT a type of supervised learning algorithm?**

a) Linear regression b) Decision trees

c) K-means clustering d) Support Vector Machines (SVM)

**32. What is the primary objective of outlier detection in Machine Learning?**

a) To identify anomalies in the data b) To preprocess the data

c) To visualize the datad) To train the model

33. **Which of the following is a classification algorithm?**

a) K-means clustering b) SVM

c) Linear regression d) Principal Component Analysis (PCA)

**34. Which evaluation metric is commonly used for regression problems?**

a) Accuracy b) Precision c) Recall d) Mean Squared Error (MSE)

**35.On which factor does my reshape depend ?**

a) No. of element present in my matrix b) loss of continuity of elements in matrix

c) No. of element present in it’s transpose matrix d) Determinant of the matrix