**Section – A (20 questions)**

**1. What does the term "overfitting" refer to in machine learning?**

a) The model performs well on training data but poorly on unseen data.

b) The model fails to capture the underlying patterns in the data.

c) The model has too few parameters to represent the complexity of the data.

d) The model is too simple to capture the relationships in the data.

**2. Which of the following is a characteristic of supervised learning?**

a) The model learns from unlabeled data.

b) The model makes predictions without any input data.

c) The model is trained using input-output pairs.

d) The model clusters data points based on similarities.

**3. What is the primary goal of using feature scaling in machine learning?**

a) To increase the dimensionality of the dataset.

b) To reduce the computational cost of model training.

c) To ensure that all features contribute equally to the learning process.

d) To introduce non-linearity into the model.

4**. Which of the following evaluation metrics is used to assess the performance of a regression model?**

a) Accuracy b) Precision c) R-squared d) F1-score

**5. What is the primary purpose of using the training set in machine learning?**

a) To evaluate the performance of the model. b) To fine-tune the hyperparameters of the model.

c) To train the model parameters on labeled data. d) To test the generalization ability of the model.

**6. Which of the following techniques is used to handle missing values in a dataset?**

a) Removing the rows with missing values.

b) Replacing missing values with the mean or median of the feature.

c) Ignoring the missing values during model training. d) All of the above.

**7. What does the term "bias" refer to in the context of machine learning models?**

a) The difference between predicted and actual values.

b) The degree of flexibility of the model.

c) The error introduced by approximating a real problem.

d) The sensitivity of the model to changes in the input data.

**8. Which of the following algorithms is NOT a type of supervised learning?**

a) Decision Trees b) K-Means Clustering c) Support Vector Machines d) Linear Regression

**9. What is the primary goal of 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.

**10. Which of the following statements about ensemble learning is true?**

a) Ensemble learning combines multiple models to increase bias and variance.

b) Ensemble learning reduces the risk of overfitting compared to individual models.

c) Ensemble learning is only applicable to regression problems.

d) Ensemble learning is less robust to noisy data.

**11. What is the primary objective of dimensionality reduction techniques in machine learning?**

a) To increase the computational cost of model training.

b) To decrease the interpretability of the model.

c) To remove irrelevant features and reduce the complexity of the dataset.

d) To introduce non-linearity into the model.

**12. Which of the following is a characteristic of unsupervised learning?**

a) The model learns from labeled data. b) The model makes predictions without any input data.

c) The model is trained using input-output pairs.

d) The model clusters data points based on similarities.

**13. What does the term "precision" measure in the context of classification models?**

a) The proportion of true positive predictions out of all positive instances.

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

c) The proportion of true negative predictions out of all negative instances.

d) The balance between true positive and false positive predictions.

**14. What is the primary goal of using the test set in machine learning?**

a) To evaluate the performance of the model. b) To fine-tune the hyperparameters of the model.

c) To train the model parameters on labeled data. d) To test the generalization ability of the model.

**15. What does the term "confusion matrix" represent in the context of classification models?**

a) A matrix that summarizes the performance of a classification model.

b) A matrix that measures the correlation between predictors and the target variable.

c) A matrix that represents the distribution of data points in feature space.

d) A matrix that measures the degree of multicollinearity among independent variables.

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

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

**17. 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.

**18. 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.

**19. 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.

**20. 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.

**Section – B (15 questions )**

**21. What is the primary purpose of using the kernel trick in Support Vector Machines?**

a) To improve model convergence. b) To transform data into a higher-dimensional space.

c) To reduce the computational complexity of the model. d) To regularize the model.

**22. Which metric is used to evaluate the performance of a binary classification model when both false positives and false negatives are equally important?**

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

**23. 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.

**24. 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.

**25. 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

**26. 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.

**27. 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

**28. 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.

**29. 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.

**30. 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.

**31. 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

**32. 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.

**33. 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.

**34. 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.

**35. 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