ML Final Assessment

ΓE MARKS (/PACKAGE-COST-DETAILS/REGENERATE-MARKS?STUDENTID=20453&ASSESSMENTID=218554)

Score Obtained: 45/50 (90%)

RETAKE (/PACKAGE-COST-DETAILS/#) VIEW REPORT

1. What is Machine learning?

1/1 ATTEMPTED

- [x] The autonomous acquisition of knowledge through the use of computer program
- [ ] The autonomous acquisition of knowledge through the use of manual programs
- [ ] The selective acquisition of knowledge through the use of computer programs
- [ ] The selective acquisition of knowledge through the use of manual programs

2. Which of the following are the Applications of Machine learning ?

1/1 ATTEMPTED

- [ ] Email Filtering
- [ ] Sentiment Analysis
- [ ] Face Recognition
- [x] All of the above

3. The Father of Machine Learning is

0/1 ATTEMPTED

- [ ] Geoffrey Everest Hinton

- Geoffery Hill
- Geoffrey Chaucer
- None of the above

4. In which industry is machine learning commonly used for fraud detection and risk management? 1/1 ATTEMPTED

- Agriculture
- Finance
- Entertainment
- Education

5.
If mean, median, and mode are all equal then distribution will be? 1/1 ATTEMPTED

- Negative Skewed
- Symmetrical
- Circular
- Positive Skewed

6. What is the primary purpose of EDA in Data Science? 1/1 ATTEMPTED

- To make data more complex
- To simplify complex data

- To discover insights & patterns in data
- To visualize data

7. How can outliers in a numerical dataset be treated? 1/1 ATTEMPTED

- Ignoring them during analysis
- Replacing them with the median value
- Removing them or transforming them
- Assigning them a weight of 0

8. Which method in pandas provides a concise summary of a DataFrame, including the number of non-null values in each column? 1/1 ATTEMPTED

- df.describe()
- df.info()
- df.head()
- df.shape()

9.
How many coefficients do you need to estimate in a simple linear regression model (One independent variable)? 1/1 ATTEMPTED

- 1
- 12
-

3

2

10.

Linear Regression is an example of?

1/1 ATTEMPTED

Supervised Learning

Unsupervised Learning

Semi-Supervised Learning

Reinforcement Learning

11. Which of the following is a commonly used metric for evaluating the performance of a Linear Regression model?

1/1 ATTEMPTED

Accuracy Score

F1 Score

Mean Squared Error (MSE)

Confusion Matrix

12. What is the primary purpose of a Loss Function in Linear Regression?

1/1 ATTEMPTED

To maximize the accuracy of predictions

To minimize the difference between predicted and actual values

To calculate the mean of the target variable

- To identify outliers in the data

13.

Which one is true?

1/1 ATTEMPTED

- Ridge and Lasso regression are techniques to reduce the model complexity and prevent over-fitting which may result from simple linear regression
- Ridge regression shrinks the coefficients and it helps to reduce the model complexity and multi-collinearity.
- Lasso regression not only helps in reducing over-fitting but it can help us in feature selection
- All of the above

14.

What type of penalty is used on regression weights in Ridge regression?

1/1 ATTEMPTED

- L0
- L2
- L1
- None of the above

15.

In Ridge regression, A hyper parameter is used called '_____' that controls the weighting of the penalty to the loss function.

0/1 ATTEMPTED

- Gamma
- Alpha
- Lambda

None of above

16. With Lasso Regression the influence of the hyper parameter lambda, as lambda tends to zero the solution approaches to 1/1 ATTEMPTED

Zero

One

Linear Regression

Infinity

17. What is the main purpose of Ridge and Lasso regularization in linear regression? 1/1 ATTEMPTED

To increase the complexity of the model

To reduce the impact of outliers

To penalize large coefficients and prevent overfitting

To simplify the model by removing unnecessary features

18.
Which of the following is used where the target variable is of categorical nature? 1/1 ATTEMPTED

Logistic Regression

Knime

Keras

Linear Regression

19.

Function which performs the role of an activation function in machine learning which is used to add non-linearity in a machine learning model. Basically, the function determines which value to pass as output and what not to pass as output.

1/1 ATTEMPTED

- cost function
- signoidal function
- sigmoid function
- none of these

20.

_____ uses Harmonic Mean in place of Arithmetic Mean as it punishes the extreme values more.

1/1 ATTEMPTED

- S-measure
- F-measure
- T-measure
- None of these

21.

Decision Tree is a display of an algorithm?

1/1 ATTEMPTED

- True
- False

-

-

22.

_____ is used to calculate the homogeneity of a numerical sample in Decision Trees

0/1 ATTEMPTED

standard deviation

Variance

Mean

Mode

23.

in the case of AdaBoost, very short decision trees were used that only had a single split, called a _____?

1/1 ATTEMPTED

optimizer step

Decision stump

weak learner

bench step

24.

Unlike in AdaBoost, the incorrect result is given a higher weightage in gradient boosting.

1/1 ATTEMPTED

TRUE

FALSE

- -

- -

25.

The boosting algorithm combines a number of weak learners to form a strong learner?

1/1 ATTEMPTED

- TRUE
- FALSE
- Depends on number of samples
- Unpredictable

26.

What do you mean by a hard margin?

0/1 ATTEMPTED

- The SVM allows very low error in classification
- The SVM allows high amount of error in classification
- The SVM allows no error in classification
- none of these

27.

Closest Point to the hyper plane are support vectors?

1/1 ATTEMPTED

- False

- True
- Unpredictable
- None of these

28.

_____ calculates the harmonic mean between precision and recall, and both depend on the false positive and false negative

1/1 ATTEMPTED

- f1 score
- f2 score
- z score
- z1 score

29. In bootstrap sampling, how is a sample generated from the original dataset?

1/1 ATTEMPTED

- By taking the mean of the original dataset
- By randomly selecting observations with replacement
- By taking the median of the original dataset
- By using a linear transformation

30. What is the primary purpose of bagging?

1/1 ATTEMPTED

- To reduce the bias of a model
-

To increase the variance of a model

To improve the accuracy and robustness of a model

To simplify the model structure

31. What is Random Forest primarily used for?

1/1 ATTEMPTED

Clustering

Dimensionality Reduction

Classification and Regression

Time Series Forecasting

32. What technique does Random Forest use to create diverse trees?

1/1 ATTEMPTED

Bagging

Boosting

K-means clustering

PCA (Principal Component Analysis)

33. What is the primary purpose of the k-NN algorithm?

1/1 ATTEMPTED

Clustering data points

Predicting continuous values

Classifying data points based on similarity

- Reducing dimensionality

34. What is a common method to determine the optimal value of k in k-NN?

1/1 ATTEMPTED

- Using the mean of the dataset
- Cross-validation
- Random selection
- Heuristic guessing

35. What is the potential downside of using a very large value of k in k-NN?

1/1 ATTEMPTED

- Increased risk of overfitting
- Increased model interpretability
- Loss of important local patterns
- Decreased computational cost

36. What is the primary assumption made by the Naive Bayes algorithm?

1/1 ATTEMPTED

- Features are correlated.
- All features are independent given the class label.
- The data follows a normal distribution.
- All features have the same weight.

37. Which type of Naive Bayes would you choose for a sentiment analysis task with text data? 1/1 ATTEMPTED

- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Bernoulli Naive Bayes
- Categorical Naive Bayes

38. What is the primary purpose of Principal Component Analysis (PCA)? 1/1 ATTEMPTED

- To increase the dimensionality of the data
- To visualize high-dimensional data in lower dimensions
- To perform clustering
- to classify data points

39. Which of the following is NOT a common technique for dimensionality reduction? 1/1 ATTEMPTED

- Linear Discriminant Analysis (LDA)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Gradient Descent
- Autoencoders

40. What is the primary goal of K-means clustering? 1/1 ATTEMPTED

- To classify data points into pre-defined categories
- To minimize the sum of squared distances between data points and their assigned cluster centers
- To maximize the distance between different clusters
- To identify outliers in the dataset

41. Which of the following is a limitation of K-means clustering? ATTEMPTED 1/1

- It can handle large datasets efficiently
- It assumes spherical clusters
- It works well with any distance metric
- It is easy to implement

42. What is the main purpose of the Elbow Method in K-means clustering? 1/1 ATTEMPTED

- To visualize the clusters
- To determine the optimal number of clusters (K)
- To compute the centroids
- To evaluate the clustering performance

43. If the WCSS continues to decrease without an elbow in the plot, what might this indicate? 0/1 ATTEMPTED

- The dataset is well-clustered

- The optimal K is very high
- The clusters are too small
- There are no meaningful clusters

44. In agglomerative hierarchical clustering, what is the initial step?

1/1 ATTEMPTED

- Choosing the number of clusters
- Calculating the distance matrix
- Merging the closest clusters
- Assigning data points to their nearest cluster

45. Which linkage method considers the maximum distance between points in two clusters?

1/1 ATTEMPTED

- Single linkage
- Complete linkage
- Average linkage
- Ward's linkage

46. Which of the following best describes agglomerative hierarchical clustering?

1/1 ATTEMPTED

- It starts with a single cluster containing all data points.
- It starts with each data point as its own cluster and merges them iteratively.

- It splits a large cluster into smaller clusters.
- It requires the number of clusters to be defined before clustering.

47. In divisive hierarchical clustering, what is the initial step? 1/1 ATTEMPTED

- Merging the closest clusters
- Creating a distance matrix
- Starting with a single cluster containing all data points
- Assigning data points to clusters

48. What does "single linkage" refer to in hierarchical clustering? ATTEMPTED 1/1

- The distance between the farthest points of two clusters.
- The distance between the closest points of two clusters.
- The average distance between all points in two clusters.
- The variance within a cluster.

49. What is the main characteristic of average linkage in hierarchical clustering? 1/1 ATTEMPTED

- It uses the closest points to define the distance between clusters.
- It averages the distances between all points in the two clusters.
- It only considers the centroid of each cluster.
-

It is not commonly used in practice.

50. What is the logistic loss function also known as? 1/1 ATTEMPTED

- Mean Squared Error
- Hinge Loss
- Cross-Entropy Loss
- Euclidean Loss