SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR (AUTONOMOUS)

Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code: Machine Learning (23CS0902)	**Regulation:** R23
Course & Branch: B.Tech – CSM,CAD & CAI	**Year & Sem:** II Year & II Sem

UNIT – I

INTRODUCTION TO MACHINE LEARNING

1.	a)	Define Machine Learning What are the advantages of Machine Learning.	[CO1][L1]	5M
	b)	How do you select Machine Learning model.	[CO1][L3]	5M
2.		Classify various types of Machine Learning & explain any ML model with an example.	[CO1][L2]	10M
3.		Explain the applications of Machine Learning.	[CO1][L2]	10M
4.	a)	Explain concepts of learning by Rote & Induction with an example.	[CO1][L2]	5M
	b)	Briefly discuss various Dataset's in Machine Learning	[CO1][L2]	5M
5.	a)	Briefly explain the concepts of Reinforcement Algorithm	[CO1][L2]	5M
	b)	Discuss about various types of Data	[CO1][L1]	5M
6.		Explain different stages in Machine Learning	[CO1][L2]	10M
7.	a)	Briefly describe the procedure to handle duplicate records	[CO1][L3]	5M
	b)	Explain about feature engineering in Machine Learning	[CO1][L3]	5M
8.		What is the role of searching and learning in Machine Learning Explain Grid search & Random search with an example	[CO1][L4]	10M
9.		Explain different Data collection Methods	[CO1][L2]	10M
10.	a)	What is the need of label encoding	[CO1][L1]	2M
	b)	What is meant by supervised and Unsupervised learning in ML	[CO1][L1]	2M
	c)	Define machine learning.	[CO1][L1]	2M
	d)	What is meant by Label Dataset in ML	[CO1][L1]	2M
	e)	What is meant by regression problem	[CO1][L1]	2M

UNIT – II

NEAREST NEIGHBOR-BASED MODELS

1.	a)	Explain the steps involved in Nearest Neighbour Models.	[CO2][L2]	5M
	b)	List out the applications of Manhattan Distance in ML.	[CO2][L1]	5M
2.		Explain Euclidian Distance measure with one example.	[CO2][L2]	10M
3.		Explain Manhattan Distance measure with example.	[CO2][L2]	10M
4.	a)	Explain the concept of Murkowski distance.	[CO2][L2]	5M
	b)	Explain hamming distance for any two binary and DNA patterns.	[CO2][L2]	5M
5.		Explain Non Metric Proximity measures with one example.	[CO2][L2]	10M
6.		Explain the steps involved in K-Nearest Neighbors algorithm	[CO2][L2]	10M
7.		Classify a new customer as "high-spending" or "low-spending" based on their age and income by using KNN algorithm.	[CO2][L3]	10M

Customer	Age	Income (Rupees)	Spending Class
A	25	30,000	Low
B	45	80,000	High
C	35	50,000	Low
D	50	90,000	High
E (New)	40	60,000	

8.		A bank wants to classify customers as "Low-Risk" or "High-Risk" for loans. Classify Customer E using Manhattan Distance and Radius Distance Nearest Neighbour Algorithm. Assume suitable radius.	[CO2][L3]	10M

Customer	Age	Income ($)	Credit Score	Loan Amount (Rupees)	Risk Level
A	25	30,000	700	10,000	Low
B	45	80,000	600	40,000	High
C	35	50,000	750	20,000	Low
D	55	90,000	580	50,000	High
E (New)	40	60,000	680	25,000	

9.		Explain KNN Regression with one example. Also list out advantages	[CO2][L2]	10M
10.	a)	List out the performance measures of Regression.	[CO2][L1]	2M
	b)	List out the performance measures of Classification.	[CO2][L1]	2M
	c)	Define MAE and R^2.	[CO2][L1]	2M
	d)	How does KNN regression differ from KNN classification	[CO2][L2]	2M
	e)	List out common distance measures used in machine learning.	[CO2][L1]	2M

UNIT – III

MODELS BASED ON DECISION TREES & THE BAYES CLASSIFIER

1.	Explain the key steps involved in building a Decision Tree for classification.	**[CO4][L2]**	**10M**
2.	Discuss how impurity is measured in decision trees using Gini Index and Entropy with examples.	**[CO3][L4]**	**10M**
3.	Describe the steps involved in Decision Tree is built for regression with one example.	**[CO3][L3]**	**10M**
4.	Describe the steps involved in training a Random Forest for classification.	**[CO4][L2]**	**10M**
5.	Apply Random Forest Regression to a sample dataset and explain how it predicts continuous target values. Mention any evaluation metrics used.	**[CO3][L3]**	**10M**
6.	Explain the key concepts of Bayes' Rule and how it is used in classification. Provide an example.	**[CO4][L2]**	**10M**
7.	Describe the steps involved in building a Bayes Classifier for binary classification using Bayes' Rule.	**[CO4][L3]**	**10M**
8.	What is the Naive Bayes Classifier Explain the assumption of class conditional independence and how it simplifies computation.	**[CO4][L2]**	**10M**
9.	Apply the Naive Bayes Classifier to a real-world multi-class problem such as spam detection. Outline the model building, prediction, and evaluation steps.	**[CO4][L3]**	**10M**
10. a)	What is the main criterion used to split nodes in a decision tree during classification	**[CO4][L1]**	**2M**
b)	State one key difference between decision trees used for classification and for regression.	**[CO3][L2]**	**2M**
c)	What is the role of the bias–variance trade-off in decision trees	**[CO3][L4]**	**2M**
d)	In the context of Bayes' Rule, what is meant by the 'prior probability'	**[CO3][L1]**	**2M**
e)	What assumption does the Naive Bayes Classifier make about features	**[CO4][L1]**	**2M**

UNIT – IV

LINEAR DISCRIMINANTS FOR MACHINE LEARNING

1.	Explain the steps involved in training a Perceptron classifier with a simple example.	**[CO5][L3]**	**10M**
2.	What is Linear Discriminant Analysis (LDA) Explain LDA steps for classification.	**[CO5][L2]**	**10M**
3.	Explain the Perceptron Learning Algorithm with steps. How does the algorithm converge for linearly separable data	**[CO5][L3]**	**10M**
4.	Describe the working of Support Vector Machines (SVMs) for linearly separable data. Explain margin, hyper plane, and support vectors.	**[CO5][L2]**	**10M**
5.	What happens when data is not linearly separable in SVM Explain how soft margin and Kernel Trick help.	**[CO5][L4]**	**10M**
6.	Explain the concept of the Kernel Trick in SVM with an example. Why is it useful for non-linear classification	**[CO5][L3]**	**10M**
7.	Compare and contrast Logistic Regression and Linear Regression. Give one example each.	**[CO5][L4]**	**10M**
8.	What is a Multi-Layer Perceptron (MLP) Describe its architecture and how it extends the single-layer perceptron.	**[CO5][L2]**	**10M**
9.	Explain the Backpropagation algorithm for training an MLP. Include forward pass, error calculation, and weight update steps.	**[CO5][L3]**	**10M**
10. a)	What is the main objective of the Perceptron Learning Algorithm	**[CO5][L1]**	**2M**
b)	Define the term 'Support Vector' in Support Vector Machines (SVM).	**[CO5][L1]**	**2M**
c)	What is the role of the activation function in a Multi-Layer Perceptron	**[CO5][L2]**	**2M**
d)	Mention one key difference between Linear Regression and Logistic Regression.	**[CO5][L1]**	**2M**
e)	What does the Kernel Trick do in the context of SVM	**[CO5][L2]**	**2M**

UNIT – V

CLUSTERING

1.	What is clustering Explain the types of clustering methods with simple examples.	[CO6][L2]	10M
2.	Describe the K-Means clustering algorithm. What are the steps involved in the algorithm Provide an example.	[CO6][L3]	10M
3.	What is Fuzzy C-Means clustering How does it differ from K-Means clustering	[CO6][L3]	10M
4.	Explain Agglomerative and Divisive hierarchical clustering methods. How are they different	[CO6][L3]	10M
5.	What is Soft Clustering How is it used in clustering methods like Fuzzy C-Means	[CO6][L4]	10M
6.	What is Matrix Factorization in clustering How does it help in clustering large datasets	[CO6][L3]	10M
7.	Explain Rough Clustering and the Rough K-Means algorithm. How is it different from traditional K-Means	[CO6][L4]	10M
8.	What is the Expectation Maximization (EM) algorithm How does it work for clustering data	[CO6][L3]	10M
9.	Explain Spectral Clustering. How does it work, and when is it useful for clustering	[CO6][L4]	10M
10. a)	What is the difference between Agglomerative and Divisive clustering methods	[CO6][L1]	2M
b)	What is a centroid in K-Means clustering	[CO6][L1]	2M
c)	What is Soft Clustering Give an example.	[CO6][L2]	2M
d)	How does Fuzzy C-Means differ from K-Means	[CO6][L2]	2M
e)	What is Matrix Factorization in clustering Why is it useful	[CO6][L2]	2M

UNIT-I

Introduction to Machine Learning

1   Machine Learning is an application of   .   [   ]
     A)Block Chain   B) Neural Network   **C) Artificial Intelligence**   D) None
2   Application of Machine Learning is.   [   ]
     A)Email filtering   B) Sentimental Analysis   C) Face Recognition   **D) All of the above**
3   Among the following option identify the one which is not a type of learning   [   ]
     A)Semi Supervised   B) Supervised C) Un Supervised   **D) Reinforcement**
4   If the coefficient of determination is a positive value, then the regression equation   [   ]
     A)must have a positive slope   B)must have a negative slope
     C)either a positive or a negative slope   D)None
5   Type of matrix decomposition model is   [   ]
     A)Predictive Model   **B) Descriptive Model**   C) Logical Model   D) Theorem Model
6   Identify the kind of learning algorithm for "facial identities for facial expressions".   [   ]
     A)Prediction   **B) Recognition patterns**   C)Recognizing anomalies   D) Generating Patterns
7   Machine learning is a subset of which of the following.   [   ]
     A)Deep Learning   B) Data Learning   **C) Artificial Intelligence**   D) Negative Function
8   Which of the following machine learning techniques helps in detecting the outliers in data[   ]
     A)Classification   B) Clustering **C) Anomaly Detection**   D)Data Set
9   The most significant phase in genetic algorithm is   [   ]
     A)Mutation   B) Selection   C) Fitness Function   **D)Crossover**
10   Which of the following are common classes of problems in machine learning   [   ]
     A)Regression B) Classification   C) Clustering **D) All of the above**
11   Which of the following is not a supervised learning   [   ]
     A)PCA   B) Naïve Bayesian   C) Linear Regression D) Decision Tree
12   Which of the following is not a machine learning algorithm   [   ]
     A)SVM   **B) SVG**   C)Random Forest   D) KNN
13   Among the following option identify the one which is used to create the most common   [   ]
     A)plot B) quickplot   **C) qplot**   D) quickerplot
14   The Real-world machine learning use cases are   [   ]
     A)Digital Assistant   B) Chatbots   C)Fraud Detection   **D) All of the above**
15   An un supervised learning problems can be grouped as   [   ]
     A)Clustering B) Association   **C) Both A & B**   D) Decision Tree
16   In Machine learning the model that must solve the given performance task is known as….[   ]
     A)Critic   B) Generalize **C) Performance System**   D) All of the above
17   Linear regression is a regression techniques that is used to model data having a outcome   [   ]
     A)Linear, Binary   **B) Linear, Numeric**   C) Nonlinear, binary D) Nonlinear, Numeric
18   In what type of learning labelled training data is used   [   ]

A)Un supervised   B) Reinforcement   C) Active learning   **D)Supervised**

19   What does dimensionality reduction reduce   [   ]

A)stochastics **B) collinerity** C) performance   D) entropy

20   Type of matrix decomposition model is   [   ]

A)Descriptive model **B) Predictive Model** C) Logical Model   D) Validation data

21   PCA is   [   ]

A)Forward features   B) Backward features **C) Feature extraction**   A) D)Feature Prediction

22   The average squared difference between classifier predicted output and actual output   [   ]

A) mean squared error B) root mean squared error C) mean absolute error D) mean relative error

23   The problem of finding hidden structure in unlabelled data is called   [   ]

A) Unsupervised   B)Supervised C) Reinforcement   D) A&B

24   Database query is used to uncover this type of knowledge.   [   ]

A) Hidden   B) Shallow   C) Deep   **D) Multi dimension**

25   Which learning Requires self-assessment to identify patterns within data   [   ]

A) Supervised   **B) UnSupervised**   C) Semi supervised   B) D) Reinforcement

26   Different learning model does not include   [   ]

A) Memorize B) Analogy   C) Deduction **D) Introduction**

27   For box plot, the upper and lower whisker length depends on   [   ]

A) Median   B) Mean   **C) IQR**   D) All of the above

28   Structure representation of raw input data into meaningful   is called a mode [   ]

A) Pattern   B) Data   C) Object   D) Set of data

29   Feature selection tries to eliminate feature which are   [   ]

A) Rich   B) Redundant **C) Irrelevant** D) B & C

30   For   Supervised   learning model.we   have   [   ]

A)Interactive **B)Predictive** C)Descriptive D)Prescriptive

31   The K- mean algorithm is a   [   ]

A)Supervised **B) Un Supervised**   C) Reinforcement   D) Weakly Supervised

32   For Unsupervised learning we have   model   [   ]

A)Interactive B) catch, finally   **C) Descriptive**   D)Prescriptive

33   Predicting whether a tumor is malignant or benign is an example   [   ]

A)Unsupervised   B) Supervised regression   **C) Supervised** D) Categorical Classification

34   Which of the following measures is not used for a classification model   [   ]

A) Accuracy   B) Precision   C) Recall   **D) Purity**

35. Simplest form of n-grams where n=1 is called   [   ]

A) Bigram **B) Unigram** C) Trigram D) none of the above

36   Cosine similarity is most popular used in   [   ]

A)Image Classification   **B) Text**   C) Feature Selection D) None

37   Different learning methods does not include   [   ]

A)Analogy   **B)Introduction**   C) Deduction D) Memorization

38   Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging   [   ]

A)Decision Tree   B) Regression C)Classification   **D)Random Forest**

39   Which of the following techniques cannot be used for normalization in text mining   [   ]

A)Stemming B) Lemmatization   **C) Stop word Removal**   D)None

40   As the number of training examples goes to infinity, your model trained on that data will have:

[   ]

A)Lower variance   **B) Higher variance** C) Same variance   D) None of the above

Unit II
Nearest Neighboring Neural Network

1. What is the primary purpose of proximity measures in data analysis   [   ]
   A) To calculate distances between data points
   B) To visualize data distributions
   C) To identify patterns in data
   D) To classify data into categories

2. What is the name of the proximity measure that calculates the straight-line distance between two points   [   ]
   A) Euclidean distance   B) Manhattan distance
   C) Minkowski distance   D) Cosine similarity

3. Which proximity measure is used to calculate the similarity between two vectors   [   ]
   A) Cosine similarity   B) Jaccard similarity
   C) Euclidean distance   D) Manhattan distance

4. What is the name of the proximity measure that calculates the distance between two points in a grid-based system   [   ]
   A) Manhattan distance   B) Euclidean distance
   C) Minkowski distance   D) Chebyshev distance

5. What is the name of the proximity measure that calculates the similarity between two sets [   ]
   A) Jaccard similarity   B) Cosine similarity
   C) Euclidean distance   D) Manhattan distance

6. Which proximity measure is used to calculate the distance between two points in a high-dimensional space   [   ]
   A) Euclidean distance   B) Manhattan distance
   C) Minkowski distance   **D) Mahalanobis distance**

7. What is the name of the proximity measure that calculates the distance between two points based on their ranks   [   ]
   A) **Spearman distance**   B) Kendall distance
   C) Euclidean distance   D) Manhattan distance

8. What is a non-metric similarity function   [   ]
   A) A function that satisfies the triangle inequality
   B) **A function that does not satisfy the triangle inequality**
   C) A function that is used for clustering
   D) A function that is used for classification

9. Which of the following is an example of a non-metric similarity function   [   ]
   A) Euclidean distance   B) Cosine similarity
   C) Jaccard similarity   **D) Edit distance**

10. What is the primary characteristic of a non-metric similarity function   [   ]
    A) It is symmetric   B) It is non-negative
    C) It does not satisfy the triangle inequality   D) It is used for clustering

11. Which of the following non-metric similarity functions is used for comparing strings   [   ]
    A) Edit distance   B) Cosine similarity   C) Jaccard similarity D) Euclidean distance

12. What is the name of the non-metric similarity function that is used for comparing sets [   ]
    A) Jaccard similarity   B) Cosine similarity
    C) Euclidean distance   D) Manhattan distance

13. Which of the following non-metric similarity functions is used for comparing graphs   [   ]

    A) Graph edit distance   B) Cosine similarity

    C) Jaccard similarity   D) Euclidean distance

14. What is the primary advantage of using non-metric similarity functions   [   ]

    A) They are more efficient to compute

    B) They are more accurate for certain types of data

    C) They are more interpretable

    D) They are more robust to outliers

15. Which of the following is a limitation of non-metric similarity functions   [   ]

    A) They are not symmetric   B) They are not non-negative

    C) They do not satisfy the triangle inequality   D) They are not suitable for clustering

16. What is the name of the non-metric similarity function that is used for comparing images [   ]

    A) Image edit distance   B) Cosine similarity

    C) Jaccard similarity   D) Euclidean distance

17. What is the primary purpose of measuring proximity between binary patterns   [   ]

    A) To classify data into categories   B) To cluster data into groups

    C) To identify similarities between patterns   D) To predict continuous outcomes

18. Which of the following proximity measures is commonly used for binary patterns   [   ]

    A) Euclidean distance   B) Manhattan distance

    C) Jaccard similarity   D) Cosine similarity

19. What is the name of the proximity measure that calculates the number of identical bits between two binary patterns   [   ]

    A) Hamming distance   B) Jaccard similarity

    C) Cosine similarity   D) Euclidean distance

20. Which of the following proximity measures is sensitive to the presence or absence of a feature in a binary pattern   [   ]

    A) Jaccard similarity   B) Cosine similarity

    C) Euclidean distance   D) Manhattan distance

21. What is the name of the proximity measure that calculates the similarity between two binary patterns based on the number of shared features   [   ]

    A) Jaccard similarity   B) Cosine similarity

    C) Euclidean distance   D) Simple matching coefficient

22. Which classification algorithm uses Euclidean distance to classify new instances   [   ]

    A) K-Nearest Neighbors (KNN)   B) Support Vector Machines (SVM)

    C) Decision Trees   D) Random Forest

23. What is the primary advantage of using K-Nearest Neighbors (KNN) algorithm   [   ]

    A) It can handle high-dimensional data   B) It is robust to outliers

    C) It is simple to implement   D) It can handle non-linear relationship

24 What is the name of the classification algorithm that uses Mahalanobis distance   [   ]

    A) Quadratic Discriminant Analysis (QDA)   B) Linear Discriminant Analysis (LDA)

    C) K-Nearest Neighbors (KNN)   D) Support Vector Machines (SVM)

25. Which classification algorithm uses cosine similarity to classify text documents   [   ]

    A) K-Nearest Neighbors (KNN)   B) Support Vector Machines (SVM)

    C) Naive Bayes   **D) Rocchio algorithm**

26. Which distance measure is commonly used in Support Vector Machines algorithm   [   ]

    A) Euclidean distance   B) Manhattan distance

    C) Minkowski distance   **D) Kernel distance**

27. What is the name of the classification algorithm that uses Manhattan distance [   ]

    A) K-Nearest Neighbors (KNN)   B) Support Vector Machines (SVM)

    C) Decision Trees   D) Random Forest

28. Which classification algorithm uses Minkowski distance   [   ]

    A) K-Nearest Neighbors (KNN)   B) Support Vector Machines (SVM)

    C) Decision Trees   D) Random Forest

29. What is the primary disadvantage of using distance-based classification algorithms   [   ]

    A) They are sensitive to outliers   B) They are computationally expensive

    C) They are not suitable for high-dimensional data   **D) All of the above**

30. What is the primary advantage of using K-Nearest Neighbor (KNN) classifiers   [   ]

    A) They can handle high-dimensional data   B) They are robust to outliers

    C) They are simple to implement   D) They can handle non-linear relationship

31. Which of the following is a characteristic of K-Nearest Neighbor (KNN) classifiers [ ]

    A) They are parametric models   **B) They are non-parametric models**

    C) They are suitable for high-dimensional data   D) They are robust to outliers

32. How does the larger value of K affect the performance of K-Nearest Neighbor (KNN) classifiers

                                                            [   ]

    A) leads to overfitting   **B) leads to underfitting**

    C) leads to befitting   D) none of the above

33. Which of the following is a limitation of K-Nearest Neighbor (KNN) classifiers   [   ]

    A) They are not suitable for high-dimensional data   B) They are sensitive to outliers

    C) They are computationally expensive   **D) All of the above**

34. What is the name of the technique used to select the optimal value of K in K-Nearest Neighbor (KNN) classifiers   [   ]

    A) Cross-validation B) Bootstrapping   C) Grid search   D) Random search

35. What is the primary goal of K-Nearest Neighbor (KNN) regression   [   ]

    A) To classify data into categories   **B) To predict continuous outcomes**

    C) To cluster data into groups   D) To identify patterns in data

36. Which of the following is a characteristic of K-Nearest Neighbor (KNN) regression   [   ]

    A) It is a parametric model   **B) It is a non-parametric model**

    C) It is suitable for high-dimensional data   D) It is robust to outlier

37. How does the value of K affect the performance of K-Nearest Neighbor (KNN) regression[   ]

    A) A small value of K leads to overfitting B) A large value of K leads to underfitting

    C) A small value of K leads to underfitting   D) A large value of K leads to overfitting

38 What is the name of the metric that measures the proportion of variance in the dependent variable that is predictable from the independent variable(s) in a regression model   [   ]

    A) R-squared   B) Coefficient of Determination

    C) Mean Squared Error (MSE)   D) Mean Absolute Error (MAE)

39.. Which of the following metrics is sensitive to outliers in a regression model   [   ]

    A) Mean Squared Error (MSE)   B) Mean Absolute Error (MAE)

    C) Root Mean Squared Error (RMSE) D) All of the above

40. What is the name of the technique used to evaluate the performance of a regression model on unseen data   [   ]

    A) Cross-validation   B) Bootstrapping

    C) Holdout method   **D) All of the above**

Unit III

Decision Trees and Bayes probability

1. What is a decision tree?   [   ]
A) type of neural network   B) **tree-like model used for classification & regression**
C) type of clustering algorithm   D) type of dimensionality reduction technique

2. Which of the following is a characteristic of decision trees?   [   ]
A) They are parametric models   **B) They are non-parametric models**
C) They are suitable for high-dimensional data D) They are robust to outliers

3. What is the primary advantage of using decision trees?   [   ]
A) They can handle non-linear relationships   B) They are robust to outliers
C) A & B   D) None of the above

4. Which of the following is a type of decision tree algorithm?   [   ]
A) ID3   B) C4.5   C) CART   **D) All of the above**

5. What is the name of the process used to reduce the size of a decision tree?   [   ]
A) Pruning   B) Trimming C) Reducing   D) Simplifying

6. Which of the following is a benefit of using decision trees?   [   ]
A) They can handle missing values   B) They are suitable for large datasets
C) They are easy to visualize   **D) All of the above**

7. What is the primary disadvantage of using decision trees?   [   ]
A) They can be prone to over fitting   B) They are sensitive to outliers
C) They are computationally inexpensive   D) All of the above

8. Which of the following is a type of split used in decision trees?   [   ]
A) Binary split   B) Multiway split   **C) Both A and B**   D) Neither A nor B

9. What is the name of the measure used to evaluate the goodness of a split in a decision tree?   [   ]
A) Gini impurity   B) Entropy   C) Information gain   **D) All of the above**

10. What is the primary purpose of impurity measures in decision tree construction?
                            [   ]
A) To evaluate the performance of the model   **B) To select the best split for a node**
C) To prune the tree   D) To handle missing values

11. Which is the better method for fast computational ?   [   ]
A) Gini impurity   B) Entropy   C) Variance   D) All of the above

12. What is Gini impurity?   [   ]
A) A measure of the variance in a dataset   B) A measure of the entropy in a dataset
C) A measure of the probability of misclassifying a sample
D) A measure of the proportion of samples in a node that belong to a single class

13. Which impurity measure is based on the concept of information theory?   [   ]
A) Gini impurity   **B) Entropy**   C) Variance   D) Mean squared error

14. What is the primary advantage of using impurity measures in decision tree construction?   [   ]
A) They help to reduce overfitting   **B) They help to select the best split for a node**
C) They help to handle missing values
D) They help to improve the interpretability of the model

15. What is the primary goal of regression based on decision trees?   [   ]
A) To classify data into categories   B) To predict continuous outcomes
C) To cluster data into groups   D) To identify patterns in data

16. Which of the following is a type of decision tree used for regression?   [   ]

A) Classification tree **B) Regression tree**   C) Decision stump   D) Random forest

17. How do regression trees handle continuous target variables?   [   ]

A) By predicting the class label **B) By predicting the mean value of the target variable**

C) By predicting the median value of the target variable

D) By predicting the mode value of the target variable

18. Which is not a common method used to improve the performance of regression trees?   [   ]

A) Pruning   **B).Clustering**   C) Bagging   D) Boosting

19. What is a random forest?   [   ]

A) A single decision tree   **B) An ensemble of decision trees**

C) A type of neural network   D) A clustering algorithm

20. Which of the following is a benefit of using random forests?   [   ]

A) Improved accuracy   B) Reduced overfitting

C) Increased disorder   D) Extrapolate well

21. How do random forests handle missing values?   [   ]

A) By ignoring them   B) By imputing them

C) By using surrogate splits   **D) By using a combination of the above**

22. What is the primary advantage of using random forests for classification and regression?   [   ]

A) They can handle high-dimensional data   B) Prone to over fitting

C) Computationally expensive   D) Wrong continuous nature prediction

23. What is Bayes' theorem used for in machine learning?   [   ]

A) To classify data into categories   B) To predict continuous outcomes

C) To cluster data into groups   **D) To update probabilities based on new evidence**

24 Which of the following is a type of Bayes classifier?   [   ]

A) Naive Bayes   B) Bayesian network **C) Both A and B**   D) Neither A nor B

25. What is the primary assumption made by the Naive Bayes classifier?   [   ]

A) Features are independent   B) Features are dependent

C) Features are normally distributed   D) Features are not normally distribute

26. Which of the following is a benefit of using Bayes classifiers?   [   ]

A) They can handle high-dimensional data   B) They are robust to outliers

C) They can handle missing values   **D) They are simple to implement**

27. What is the Bayes decision rule?   [   ]

A) Choose the class with the highest prior probability

B) Choose the class with the highest posterior probability

C) Choose the class with the lowest prior probability

D) Choose the class with the lowest posterior probability

28. Which of the following is a type of probability used in Bayes' theorem?   [   ]

A) Prior probability   B) Likelihood   **C) A & B**   D) None of the above

29. What is the primary advantage of using Bayes classifiers?   [   ]

A) They can handle non-linear relationships B) They are robust to outliers

C) They can handle missing values **D) They are based on sound mathematical principles**

30. Which of the following is an application of Bayes classifiers?   [   ]

A) Text classification   B) Image classification   **C) A & B**   D) None

31. Which of the following is a limitation of Bayes classifiers?   [   ]

A) They assume independence of features B) They are sensitive to outliers

C) They are computationally expensive   D) All of the above

32. What is the primary purpose of Bayes' rule in inference?   [   ]

A) To calculate prior probabilities **B) To update probabilities based on new evidence**

C) To determine the likelihood of an event   D) To calculate posterior probabilities

33. What is the posterior probability in Bayes' rule?   [   ]
A) The probability of the hypothesis before observing the data
B) The probability of the data given the hypothesis
C) The probability of the hypothesis after observing the data
D) The probability of the data

34. Which of the following best describes the process of Bayesian inference?   [   ]
A) Updating prior beliefs with new data to form posterior beliefs
B) Using prior beliefs to predict future outcomes
C) Ignoring prior beliefs and relying solely on new data
D) Using posterior beliefs to determine prior probability

35.What is a key advantage of Bayesian inference?   [   ]
A) It provides a fixed and certain answer
B) It allows for the incorporation of prior knowledge
C) It is only suitable for simple problems
D) It ignores uncertainty in the data

36. What is the Bayes classifier considered in terms of classification performance? [ ]
A) A heuristic approach   **B) An optimal classifier**
C) A suboptimal classifier   D) A random classifier

37. Which of the following is a property of the Bayes classifier?   [   ]
A) It minimizes the expected risk   B) It maximizes the expected risk
C) It ignores prior probabilities   D) It relies solely on likelihood

38. What is the primary reason why the Bayes classifier is considered optimal? [   ]
A) It uses a complex decision boundary   B) It minimizes the Bayes risk
C) It ignores uncertainty in the data   D) It relies on a specific distribution

39. Which of the following statements about the Bayes classifier is true?   [   ]
A) It is only suitable for binary classification
B) It requires a large amount of training data
C) It provides the lowest possible error rate
D) It is computationally expensive

40. What is the Bayes error rate?   [   ]
A) The minimum possible error rate for a classifier
B) The maximum possible error rate for a classifier
C) The average error rate for a classifier
D) The error rate for a specific classifier