Shivaji University, Kolhapur
Question Bank for Mar 2022 (Summer) Examination

Subject Code: 81549 Subject Name: Machine Learning

MCQs

1. What is machine learning?

A) Machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to evolve behaviours based on empirical data, such as from sensor data or databases.

B) "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

C)"A branch of artificial intelligence in which a computer generates rules underlying or based on raw data that has been fed into it."

D) All of above.

2. Which ML algorithm is suitable when we want to predict any continuous value?

A) Classification.

B) Regression.

C) Clustering.

D) None of the above

3. Cleaning of Data is done in _____

A) Data Collection

B) Data Preparation.

C) Data Splitting.

D) Data Testing.

4. Which of these are classification tasks?

A) Find the gender of a person by analyzing his writing style.

B) Predict whether there will be abnormally heavy rainfall next year.

C) Both A & B.

D) None of above.

5. In the regression equation y = b0 + b1x, b0 is the

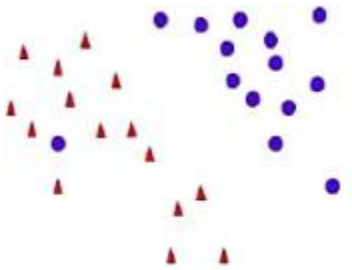
A) slope of the line

B) independent variable

C) y intercept

D) none of the above

6. What might be the best complexity of the curve which can be utilized for isolating the two classes displayed in the picture down?



A) Linear

B) Quadratic

C) Cubic

D) Insufficient data to draw conclusion

7. Suitable evaluation metric for measuring the performance of a given regression model is…

A) Mean absolute error

B) Root mean square error

C) Both A and B

D) None of above

8. What type of machine learning is suitable for predicting the dependent variables with two different values?

A) Logistic Regression

B) Linear Regression

C) Multiple linear Regression

D) Polynomial Regression

9. Which of following are categorical features?

A) Height of a person

B) Price of petroleum

C) Mother tongue of a person

D) Amount of rainfall in a day

10. Let's say in our target marketing problem, we work on 10,000 customer records to predict which customers are likely to respond to our marketing effort. Considering the below observation calculate the Recall?

Action	Predicted (that there will be a buy)	Predicted (that there will be no buy)
Actually bought	TP: 500	FN: 400
Actually did not buy	FP: 100	TN: 9000

A) 95%

B) 83.33%

C) 55.55%

D) 40%

11. Appropriate chart for visualizing the linear relationship between two variables is….

A) Scatter plot

B) Bar Chart

C) Histogram

D) None of the above.

12. ___gives the rate of speed where the gradient moves during gradient descent.

A) Learning rate

B) Cost Function

C) Hypothesis Function

D) None of above

13. What is formula to calculate error of single data point?

A) Actual value – Predicted value.

B) Actual value + Predicted value.

C) Predicted Value – Actual Value.

D) Predicted Value + Actual Value

14 --------is used to optimize the cost function or the error of the model.

A) Gradient Descent Algorithm

B) Hypothesis Function

C) Both a and b

D) None of above

15. Using gradient descent algorithm we get

A) Slope.

B) Intercept.

C) Slope and intercept.

D) Slop Intercept and error.

16. _____is a measure of how wrong the model is in terms of its ability to estimate the relationship between x and y.

A) Cost Function

B) Hypothesis Function

C) both A and B

D) None of above

17. The decision trees are most suitable for

A) For tabular data.

B) When the output required is discrete.

C) The training data may contain missing attribute values.

D) All of the above

18 ---------- is the randomness in data and metric to use impurity.

A) Information Gain

B) Gini Index

C) Variance

D) Entropy

19. Random Forest uses:

A) Ensemble Techniques

B) Bagging

C) Boosting

D) All of the above

20. Classification is what type of machine learning technique?

A) Supervised

B) Unsupervised

C) Both a and b

D) None of above

21. If we train a logistic regression model with 200 numbers of instances and accuracy is 0.8 then calculate number of failures?

A) 160

B) 40

C) 20

D) 80

22. In the given formula P(X/Y) = (P(Y/X)*P(X))/P(Y), P(X/Y) is?

A) Posterior Probability

B) Likelihood

C) Prior Probability

D) Evidence

23. If a patient has fever, what's the probability he/she has cold?

Given data:

-A doctor knows cold causes fever 50% of time.

-Prior probability of any patient having cold is 1/50000.

-Prior probability of any patient having fever is 1/20.

A) 0.2

B) 0.02

C) 0.002

D) 0.0002

24. Consider the given data set and give the prediction whether student will be Qualified or Not qualified using KNN classifier for K=1.

- **Query = Math's (5) and Computer Science (8)**

Maths	Computer Science	Status
4	4	NQ
6	8	Q
7	9	Q
5	5	NQ
7	7	Q

A) Not Qualified

B) Qualified

C) Cannot Classif**y.**

D) None of the above

25. In one vs one classifier, if there are 4 classes then_____number of binary classifiers are required

A) 6

B) 8

C) 4

D) 2

26. "The Current state of the system depends only on the previous state of the system", is property of__

A) Bayesian Classifier

B) Hidden markov model

C) Clustering

D) None of above

27. Which of the following is not an advantage of Decision Tree?

A) Decision trees generate understandable rules.

B) Decision trees perform classification without requiring much computation.

C) Decision trees are capable of handling both continuous and categorical variables.

D) Decision trees are prone to errors in classification problems with many classes and a relatively small number of training examples.

28. A root node in Decision tree is selected based on:

A) Highest information Gain

B) Lowest information gain

C) Moderate Information gain

D) None of the above

29. Pruning is

A) Removing of unwanted branches of the tree.

B) Formed by splitting of Tree

C) Dividing the root node in to different parts.

D) Roots divided into homogeneous sets

30. Which is not an advantage of SVM

A) High Memory management

B) Handles nonlinear data efficiently

C) Capable of handling outliers

D) Handles high dimensional space.

31. Support vector machine is an algorithm used for:

A) Optimal Decision boundary

B) To support the vectors

C) Linear classification

D) None of the above

32. To transform data in to higher dimension ---------is used.

A) Kernel

B) Kernel trick

C) Nonlinear Kernel

D) All of the above.

33. What type of distance matrices are used to calculate distance between two points in hierarchical clustering?

A) Euclidean distance.

B) Manhattan distance.

C) Maximum distance.

D) All of these.

34. What is adaline in neural networks?

A) Adaptive line element

B) Adaptive linear element

C) Automatic linear element

D) None of the mentioned

35. Which is true for neural networks?

A) It has set of nodes and connections

B) Each node computes its weighted input

C) Node could be in excited state or non-excited state

D) All of the above

36. Neural networks can be used in different fields. such as –

A) Classification

B) Data processing

C) Compression.

D) All of the above

37. Why are recommendation engines becoming popular?

A) Users have lesser time, more options and face an information overload

B) It is mandatory to have recommendation engine as per telecom rules

C) It is better to recommend than ask user to search on mobile phones

D) Users don't know what they want

38. What are different Recommendation Engine techniques?

A) Content based filtering

B) Collaborative filtering

C) Knowledge based system

D) All of the above

39. What are the challenges in Content Based Filtering?

A) Need to capture significant amount of users' information, which may lead to regulatoryand pricing issues

B) Need to have information of all users across different demographics

C) Need to have lower number of categories for content based filtering to be effective

D) Need to have user's social media and digital footprint

40. What kind of information does a Recommendation Engine need for effectiverecommendations?

A) Users' explicit interactions such as information about their past activity, ratings, reviews

B) Users' implicit interactions such as device they use for access, clicks on a link, location,and dates

C) Other information about profile, such as gender, age, or income levels

D) All of the above

Unit 1

1. What is machine learning? Explain types of machine learning.

2. Explain supervised learning.

3. Explain unsupervised learning.

4. Explain reinforcement learning.

5. Explain machine learning problem categories.

6. Explain supervised learning problem categories.

7. Explain unsupervised learning problem categories

8. Draw and explain machine learning architecture.

9. Draw and explain machine learning lifecycle.

10. Explain performance measures for machine learning

Unit 2

1. Explain simple linear regression.

2. Explain gradient descent for simple linear regression.

3. What is hypothesis function for simple linear regression?

4. Explain simple regression in matrix form.

5. Explain Least Squares in Matrix Form.

6. Explain Sampling Distribution of Estimators.

7. Using the given data set find the value y when x=10.

X={1,1,2,3,4,4,5,6,6,7}

Y={2.1,2.5,3.1,3.0,3.5,3.2,4.3,3.9,4.4,4.8}

8. Using the given data set find the value y when x=10.

X={1,2,3,4,5,6}

Y={25,35,42,50,55}

9. Explain multivariate linear regression.

10. What is hypothesis function for multivariate linear regression?

Unit 3

1. Explain logistic regression.

2. What is Hypothesis representation in logistic regression?

3. Explain decision boundary logistic regression.

Regularization

4. What is cost function for logistic regression?

5. Explain Gradient Descent for Logistic Regression.

6. Explain Naïve Bayes Classifier

7. What is Over fitting &Under fitting

8. Explain instance based classifier.

9. Explain K- Nearest Neighbor Classifier

10. Explain Bayesian Network

Unit 4

1. What is decision tree? State the advantages, and limitations.

2. What is the need of decision tree?

3. Explain decision tress algorithm.

4. What is information gain and entropy in decision tree?

5. Which are algorithms used in decision tree?

6. What is SVM? Explain in detail.

7. Explain Hyperplane and Support Vectors in the SVM algorithm

8. Which are the Pros and Cons of SVM Classifiers?

9. What is kernel trick in SVM?

10. What is cost function of SVM?

Unit 5

1. What is clustering? Explain in detail.

2. Explain K Means clustering.

3. Explain Elbow Method in K Means clustering

4. Explain Hierarchical clustering.

5. What is Agglomerative Hierarchical clustering?

6. Explain working of dendrogram in Hierarchical clustering.

7. Explain Association Rule mining.

8. Explain apriori algorithm.

9. Explain Eclat algorithm

10. Explain F-P Growth algorithm

Unit 6

1. What is neural network? Explain in detail.

2. What is hypothesis function and cost function for neuron?

popularity

3. Explain gradient descent for neuron.

4. Explain Multiclass classification with neural network.

5. Explain Learning in neural network-back propagation algorithm.

6. Explain Content based recommendation engines.

7. Explain Classification based recommendation engine.

8. Explain Collaborative filtering.

9. Which are applications of neural networks?

10. Explain Collaborative Filtering in recommendation system.