Subject : Machine Learning

Class : III -I SEC A&B

Department : Artificial Intelligence & Data Science

1. Application of Machine learning \_\_\_\_\_\_\_ []

1. E-mail filtering
2. Sentimental analysis
3. face recognition
4. All of the above

2. Which among the following learning technique uses labelled data \_\_\_\_\_\_ []

1. Supervised
2. unsupervised
3. Batch learning
4. Offline learning

3. among the following options identify the one which is false regarding regression []

1. It is used for prediction
2. it is used for interpretation
3. it relates inputs to outputs
4. it discovers casual relationship

4. difficulties with K-Nearest Neighbour algorithm []

1. Curse of dimensionality
2. calculate distance of test case from all training cases
3. Both a and b
4. none of the above

5. In general, to have a well-defined learning problem we must identify []

1. class of tasks
2. measure of performance to be improved
3. source of experience
4. all of the above

6. regression model in which predictions are formed from single, univariate feature of data []

1. simple regression model
2. multiple regression model
3. lasso regression
4. Logistic regression

7. Mathematical expression for absolute error calculation is []

1. Σi|yi - f(xi)|
2. Σi(yi - f(xi))2
3. Σi|yi + f(xi)|
4. Σi(f(xi)

8.Match the below options correctly- In y=w0+w1x []

I. Y (a). real-valued coefficients

II. X (b). Target

III. W0  and w1  (c). Features

1. I-c, II-b, III-a
2. I-b, II-c, III-a
3. I-a, II-b, III-c
4. I-b, II-a, III-c

9. GradientDecent loss is a []

1. second order Optimisation algorithm
2. first order Optimisation algorithm
3. worst case algorithm
4. average algorithm

10. batch gradient descent takes []

1. single Observation for updation
2. entire data set for updation
3. subset of data for updation
4. none of the above

11. Match the following []

A) Mean absolute error i) R2 =( SSres – SStotal) / SStotal

B) Mean squared error ii) Square root of (1/N ∑(yi – f(xi))2)

C) Root mean squared error iii) R2 = 1 - SSres /SStotal

D) R squared iv) (1/N) ∑i |yi – f(xi)|

v) (1/N) ∑i (yi – f(xi)|)2

1. A-iv , B-v, C-ii, D-iii
2. A-v , B-iv, C-ii, D-iii
3. A-ii , B-iii, C-v, D-iv
4. A-iii , B-ii, C-iii, D-iv

12. \_\_\_\_\_\_\_ refers to a statistical model or a machine learning algorithm in which itneither performs well on training data nor on testing data []

1. good fitting
2. overfitting
3. underfitting
4. all of the above

13. L1 regularisation can also be called as []

1. ridge regression
2. lasso regression
3. ridge-lasso
4. lasso-ridge

14. Choose the correct option []

1. L1 regularisation Robust to outliers
2. L2 regularisation is not robust to outliers
3. L1 regularisation solution is non-sparse
4. L2 regularisation solution is non-sparse

a) A,B,C b) A,B,D c) A,B d) A,C,D

15. which is not suitable for large datasets with hundreds and thousands of samples []

1. K-fold Cross validation
2. leave-one-out cross validation
3. holdout method
4. all of the above

16. example of clustering algorithm []

1. decision tree
2. random forest
3. K-means
4. gradient decent

17. purpose of confusion Matrix in Machine Learning is to \_\_\_\_ []

1. to compare performance of different model
2. to evaluate performance of classification model
3. to evaluate performance of regression model
4. none of the above

18. purpose of regularization \_\_\_\_\_\_ []

1. to prevent overfitting
2. to increase accuracy of model
3. to reduce variance of model
4. to reduce bias of model

19. which of the following is hyperparameter in machine learning? []

1. learning rate
2. training data
3. test data
4. validation state

20. difference between validation set and test set []

1. validation set used to tune hyperparameters of a model and test set used to evaluate its performance
2. both are same
3. validation set is not necessary in Machine Learning
4. test set is not necessary in Machine Learning

21. for a dataset regression equation is Y = 21 - 3x. The correlation Coefficient is []

1. zero
2. negative
3. positive
4. none of the above

22. in case of overfitting the gap between training and test error is []

1. larger
2. smaller
3. -1
4. none of the above

23. The scenario where the model has low error rate in training data but high error rate in testing data is referred as \_\_\_\_\_\_\_\_\_\_\_ []

1. underfitting
2. overfitting
3. zero fitting
4. positive fitting

24. which of the following is not a machine learning algorithm []

1. support vector machine(SVM)
2. scalable vector graphic(SVG)
3. random forest
4. none of the above

25. example of dimensionality reduction technique []

1. principal component analysis(PCA)
2. support vector machine(SVM)
3. k-nearest neighbour
4. none of the above

26. example of non-parametric machine learning algorithm []

1. linear regression
2. Logistic regression
3. decision tree
4. support vector machine

27. Which of the following is a disadvantage of using a decision tree for classification? [ ]

A. They are not able to handle categorical variables.

B. They are not able to handle continuous variables.

C. They can only classify data into a fixed set of categories.

D. They cannot handle imbalanced datasets.

28. Which of the following evaluation measures is most appropriate for imbalanced datasets? [ ]

A. Accuracy.

B. Precision.

C. Recall.

D. F1-score.

29. Which of the following evaluation measures is most appropriate for evaluating the

performance of a classifier on a large dataset? [ ]

A. Accuracy.

B. Precision.

C. Recall.

D. F1-score.

30. Which of the following evaluation measures is most appropriate for evaluating the performance of a classifier when both false positives and false negatives are equally important? [ ]

A. Accuracy.

B. Precision.

C. Recall.

D. F1-score.

31. What is the primary goal of regression analysis? [ ]

A. To identify relationships between independent and dependent variables.

B. To classify data into distinct categories.

C. To measure the central tendency of a dataset.

D. To identify the most important features in a dataset.

32. Which type of regression is used to predict a continuous dependent variable? [ ]

A. Logistic regression

B. Linear regression

C. Decision Tree regression

D. Random Forest regression

33. What is the difference between simple linear regression and multiple linear regression? [ ]

A. Simple linear regression is used for categorical data while multiple linear regression is used for numerical data

B. Simple linear regression is a classification technique while multiple linear regression is a clustering technique

C. Simple linear regression has only one independent variable while multiple linear regression has multiple independent variables

D. simple linear regression has multiple independent variables while multiple linear regression has only one independent variable

34. What is logistic regression used for? [ ]

A. To predict a continuous outcome variable

B. To predict a categorical outcome variable

C. To analyse the relationship between two continuous variables

D. None of the above

35. Which of the following is not a type of logistic regression? [ ]

A. Binary logistic regression

B. Multinomial logistic regression

C. Ordinal logistic regression

D. Linear logistic regression

36. What is the purpose of the logistic function in logistic regression? [ ]

A. To transform the dependent variable into a categorical variable

B. To transform the independent variable into a categorical variable

C. To convert the linear equation into a probability

D. None of the above

37. How is the performance of a logistic regression model evaluated? [ ]

A. By the R-squared value

B. By the mean squared error

C. By the accuracy, sensitivity, specificity, and AUC-ROC

D. None of the above

38. In k-NN, what does the value of k represent? [ ]

A. The number of nearest neighbours to be considered

B. The distance between two data points

C. The threshold value for classification

D. None of the above

39. What is the distance metric used in k-NN? [ ]

A. Euclidean distance

B. Manhattan distance

C. Minkowski distance

D. All of the above

40. What is the main disadvantage of k-NN? [ ]

A. It requires a large amount of training data

B. It is sensitive to outliers

C. It is not suitable for high-dimensional data

D. None of the above

41. Successful applications of ML

(a) Learning to recognize spoken words (b) vehicle Learning to classify new astronomical structures

(c) Learning to drive an autonomous. (d) All of the above [ ]

42.Designing a machine learning approach involves:-

a) Choosing the type of training experience b)Choosing the target function to be learned

c)Choosing a representation for the target function d)All of the above [ ]

43. In the regression equation Y = 75.65 + 0.50X, the intercept is [ ]

a) 75.65 b)0.50 c) 1.00 d) indeterminable

44 . \_\_\_\_\_processes all the training examples for each iteration of gradient descent. [ ]

a) Stochastic Gradient Descent b) Batch Gradient Descent

c)Mini Batch gradient descent d)None of the above

45. Below are the 8 actual values of target variable in the train file.

[0,0,0,1,1,1,1,1]. What is the entropy of the target variable? [ ]

a)5/8 log(5/8) + 3/8 log(3/8) b)-(5/8 log(5/8) + 3/8 log(3/8))

c)3/8 log(5/8) + 5/8 log(3/8) d)5/8 log(3/8) – 3/8 log(5/8)

46. K-NN algorithm does more computation on test time rather than train time [ ]

a)True b)False

47. Which of the following are advantages of the logistic regression? [ ]

a) Logistic Regression is very easy to understand b)It requires less training

c)It performs well for simple datasets as well as when the data set is linearly separable d)All of the above

48. Identify the difficulties with the K-Nearest Neighbor algorithm [ ]

a)Curse of dimensionality b)Calculate the distance of the test case from all training cases

c)Both A and B d)None of the above

49. In general to have a well-defined learning problem, we must identify which of the following

(a) The class of tasks (b) The measure of performance to be improved

(c) The source of experience (d) all of the above [ ]

50. Which of the following will be Euclidean Distance between the two data point A(1,3) and B(2,3)?

a) 1 b)2 c)4 d)8 [ ]

51. Which of the following regularization technique is used for feature selection [ ]

a) Lasso b) Ridge c)Both d)None of the above

52. Which of the following is a categorical data? [ ]

a ) Branch of bank b)Expenditure in rupees c)Price of house d)Weight of a person

53. A dataset with 8300 records is split into train and test sets with a test\_split ratio of 0.18. How many records will participate in training [ ]

a) 6806 b) 8300 c) 8285 d) 1494