**MCQ**

1. AI cannot do which of the following?

a) Taking a decision while running a company

b) Recommend a product for you

c) Classfiying an image

d) Detecting a suspicious movement in CCTV camera

2. Given is the dictionary as s = {'a':[23.1],'b':[1,5,2,5]}. We want to add one key element 'd' in the dictionary s. Which option is correct?

a) s{'d'=8}

b) s['d']:8

c) s['d'] = 8

d) s(d) = 8

3. We have a dataset called df containing numeric variable n1 and categorical variable c1. We want to subset the dataset selecting the data with c1='A' and n1>34. Which of the following statement are correct?

a) df[(df['c1'] == 'A') & (df['n1'] > 34)]

b) df[df['c1'] == 'A' & df['n1'] > 34]

c) df[c1' == 'A' & n1 > 34]

d) df{df['c1'] == 'A' & n1 > 34}

4. Min Max scaler transforms the data to the range

a) (1,10)

b) (-infinity, infinity)

c) (0,infinity)

d) (0,1)

5. Which of the following is a task which does not require machine learning?

a) Predicting amount of cash dispensed from ATM

b) Finding an object in the image

c) Finding compounded interest based on some set of rules

d) Predicting the bad debts

6. \_\_\_\_\_\_\_\_\_ is an unsupervised learning algorithm

a) Support Vector Machines

b) Decision Tree

c) Linear Regression

d) Clustering Algorithm

7. Basically support vector machines for classification is a binary classifier.

a) TRUE

b) FALSE

c)

d)

8. 1. support vector machines - radial a. k (n\_neighbors)

2. decision trees b. min\_samples\_split, min\_samples\_leaf, max\_depth

3. k-nearest neighbours c. learning\_rate

4. random forest d. max\_features

5. gradient boosting e. C, gamma

a) 1-e,2-b,3-a,4-d,5-c

b) 1-a,2-b,3-c,4-d,5-e

c) 1-e,2-d,3-a,4-d,5-c

d) 1-d,2-b,3-a,4-e,5-c

9. You require to predict whether the customer will buy insurance or not. Given the data, which of the functions from sklearn will you think of using?

a) SVC, KNeighborsRegressor, LinearRegression,RandomForestClassifier,DecisionTreeClassifier

b) SVC, KNeighborsClassifier, LinearRegression,RandomForestClassifier,DecisionTreeClassifier

c) SVC, KNeighborsClassifier, LogisticRegression,RandomForestRegressor,DecisionTreeClassifier

d) SVC, KNeighborsClassifier, LogisticRegression,RandomForestClassifier,DecisionTreeClassifier

10. Bigger the variation in the data, more mathematically challenging will be the predictive modelling

a) TRUE

b) FALSE

c)

d)

11. You want to find recall, precision and F1 score for a classification type of problem. We need to use \_\_\_\_\_\_\_\_ function from sklearn.

a) sklearn.metrics.confusion\_matrix

b) sklearn.metrics.classification\_report

c) sklearn.metrics.mean\_squared\_error

d) sklearn.metrics.mean\_absolute\_error

12. For a sklearn function of supervised model fitting using GridSearchCV, you have 3 parameters with each having 5, 3 and 2 distinct values respectively. For how many parameter sets will the GridSearchCV.fit( ) run?

a) 10

b) 300

c) 100

d) 30

13. We require to shuffle the data with function of Kfold for \_\_\_\_\_\_\_\_\_\_\_\_

a) having randomized order of the data

b) optimization of errors

c) reducing the data

d) None of Above

14. The best k for k-means clustering can be found by considering value of attribute \_\_\_\_\_\_\_\_\_\_\_\_ of Kmeans object.

a) cluster\_centers\_

b) inertia\_

c) labels\_

d) n\_iter\_

15. It would be better if we modify learning rate for every epoch for neural network.

a) TRUE

b) FALSE

c)

d)

16. You have been given a A.I. task of identifying the alphabets and numbers on a car number plate and accordingly output the registration number of the car. Which of the type of algorithm is it ?

a) Multi-class classification

b) Binary Classification

c) Only regression

d) Combination of classification and regression

17. Which of the tasks cannot be fulfilled with recurrent neural networks?

a) Predicting sales of next month

b) Building a language model

c) Predicting a customer being defaulter

d) Predicting the next word

18. For solving the problem of vanishing gradient descent in recurrent neural networks, which algorithm can be used?

a) Long Short Term Memory

b) One to Many

c) Many to One

d) Pooling

19. Auto-Encoder is supervised learning algorithm

a) TRUE

b) FALSE

20. Leaky Rectified Linear Unit saturates zero values for its output.

a) TRUE

b) FALSE

21. Momentum optimizer is an improvement on ADAM optimizer

a) TRUE

b) FALSE

22. Which of the following cannot be called as Corpus text?

a) Email received by you

b) Your Tweeter tweets

c) Recorded phone calls audio clips

d) Collection of word documents you have

23. Porter's algorithm is for which of the tasks?

a) Removing punctuations

b) Removing Stop words

c) Lemmatization

d) Stemming

24. More rare is the word in the corpus, bigger will be its inverse document frequency.

a) TRUE

b) FALSE

25. Skipgram and CBOW are convolutional neural network-based models

a) TRUE

b) FALSE

26. Speciality about BERT over skipgram/CBOW is that it is

a) Bidirectional

b) Recurrent network

c) Binary Encoder Representation

d) Error Reduced model

27. You are having a multi-class classification problem for classifying an image. Softmax function is correct as activation function for the last layer.

a) TRUE

b) FALSE

28. A deployment code for operationalizing any ML should be equivalent to just running predict ( ) like function.

a) TRUE

b) FALSE

29. It is better to use pre-trained model as it is.

a) TRUE

b) FALSE

30. Match the following

1. YOLO a. Pre-trained model with CNN

2. BERT b. Faster than CPU

3. Inception-V3 c. Bag of Words

4. TF-IDF d. Object detection

5. GPU & TPU e. Pre-trained NLP Model

Which of the following is perfect matched pairs?

a) 1-c,2-e,3-a,4-d,5-b

b) 1-d,2-e,3-a,4-c,5-b

c) 1-d,2-b,3-a,4-c,5-e

d) 1-d,2-e,3-b,4-c,5-a

1. Consider the following classification confusion matrix. What is the recall for category “1”?

|  |  |  |
| --- | --- | --- |
|  | Predicted 1 | Predicted 0 |
| Existing 1 | 67 | 45 |
| Existing 0 | 34 | 98 |

1. (67+98) / (67+45+34+98)
2. 67/(67+34)
3. 67/(67+45)
4. 67/(67+45+34+98)
5. You want to predict number of units defective. There are some features for this problem. This problem is of Regression type.
   1. True
   2. False
6. You want to find recall, precision and F1 score for a classification type problem. You will use which function?
   1. sklearn.metrics.confusion\_matrix
   2. sklearn.metrics.classification\_report
   3. sklearn.metrics.mean\_squared\_error
   4. sklearn.metrics.mean\_absolute\_error
7. For calculating area under the curve (AUC) in case of receiver operating characteristics curve (ROC) we require which of following as inputs to our sklearn.metrics.roc\_auc\_score function?
   1. Test set response object(1 or 0) and Test set prediction (1 or 0)
   2. Test set response object(1 or 0) and Training set response object (1 or 0)
   3. Test set response object(1 or 0) and Training set predicted probability for 1
   4. Test set response object(1 or 0) and Test set predicted probability for 1
8. In decision trees algorithm for classification, the binary split is made on data set and its subsets for reducing which of the metrics?
   1. ROC – AUC
   2. Mean Squared Error
   3. Gini’s Impurity Index
   4. F1 score
9. In decision trees algorithm for regression, the binary split is made on data set and its subsets for reducing which of the metrics?
   1. ROC – AUC
   2. Mean Squared Error
   3. Gini’s Impurity Index
   4. F1 score
10. The tuning parameters for SVM – RBF are mainly
    1. Cost, gamma
    2. Cost only
    3. Gamma only
    4. None of the above
11. The tuning parameters for SVM – Linear are mainly
    1. Cost, gamma
    2. Cost only
    3. Gamma only
    4. None of the above
12. The base learner used for Gradient Boosting algorithms is
    1. Logistic Regression
    2. Decision Tree Classifier
    3. SVM Classifier
    4. Linear Regression
13. For Computer Vision Deep Learning problem, identifying the position of an object can be regression problem also and not just classification problem.
    1. True
    2. False
14. In NLP, bigger the inverse document frequency (IDF),
    1. Rarer is the term in the corpus
    2. More frequent is the term in corpus
    3. Nothing related to frequency in corpus
    4. None of the Above
15. Consider the following classification confusion matrix. What is the precision for category “1”?

|  |  |  |
| --- | --- | --- |
|  | Predicted 1 | Predicted 0 |
| Existing 1 | 67 | 45 |
| Existing 0 | 34 | 98 |

1. (67+98) / (67+45+34+98)
2. 67/(67+34)
3. 67/(67+45)
4. 67/(67+45+34+98)
5. You have 4 folders with images of 4 different birds namely, Peacock, Macaw, Penguin and Kingfisher. You intend to build a CNN model for classifying images of these birds. Irrespective of the convolutional layers and dense layers you have, which one of the statements written below are correct for this problem of Image classification?
   1. model.compile(optimizer=RMSprop(lr=0.001), loss='binary\_crossentropy', metrics = ['acc'])
   2. model.compile(optimizer='sgd',loss='mean\_squared\_error',metrics=['mae'])
   3. model.compile(optimizer='adam', loss='binary\_crossentropy',metrics=['accuracy'])
   4. model.compile(optimizer=RMSprop(lr=0.001), loss='categorical\_crossentropy', metrics = ['acc'])
6. You have a regression problem for estimating the number of units sold. You intend to build a MLP model for this problem. Irrespective of the number of dense layers and other options related to dense layers, which one of the statements written below are correct for this problem of prediction?
   1. model.compile(optimizer=RMSprop(lr=0.001), loss='binary\_crossentropy', metrics = ['acc'])
   2. model.compile(optimizer='sgd',loss='mean\_squared\_error',metrics=['mae'])
   3. model.compile(optimizer='adam', loss='binary\_crossentropy',metrics=['accuracy'])
   4. model.compile(optimizer=RMSprop(lr=0.001), loss='categorical\_crossentropy', metrics = ['acc'])
7. For a time series numeric prediction problem, which one of the algorithms do you find best suited?
   1. Long Short Term Memory (LSTM) Model
   2. Decision Tree Classifier
   3. Linear Regression
   4. Logistic Regression
8. Which one of the following algorithms is unsupervised learning algorithm?
   1. Convolutional Neural Network
   2. Decision Tree Regression
   3. Auto-Encoders
   4. Multi-Layer Perceptron (Fully Connected Layers)
9. There are three models, you have tried to fit to a regression type of problem. One has R-Squared (r2\_score) as 0.56, second one has R-squared as 0.44 and third one has R-squared as 0.78. Which of the models is better?
   1. Model 1
   2. Model 2
   3. Model 3
   4. None of them, as R-Squared is not supposed to be a metric for this type of problem
10. There are three models, you have tried to fit to a classification type of problem. One has R-Squared (r2\_score) as 0.56 and second one has R-squared as 0.44 and third one has R-squared as 0.78. Which of the models is better?
    1. Model 1
    2. Model 2
    3. Model 3
    4. None of them, as R-Squared is not supposed to be a metric for this type of problem
11. You intend to scale the values in your training dataset in such a way that every scaled value fall in the range between 0 to 1. Which scaler will you apply on the data?
    1. Min Max Scaler
    2. Standard Scaler
    3. Normalized Scaler
    4. Just a Bucketizer to be applied
12. Given a corpus, what will be your first task among the activities mentioned below?
    1. Tokenizing the text
    2. Cleaning the text
    3. Transforming every word in the corpus to numerical vector
    4. Applying any Machine Learning Algorithm