Score: 94%

Congratulations! You have now passed the final test of Machine Learning training.

Download Certificate

Final Test Summary

No. of questions: 35 | Correct answers: 33 | Incorrect answers: 2

Show incorrect attempt only

Question 1

1 Mark

On what basis you can differentiate between a classifications and a regression problem?

A	The dependent variable	✓
B	The independent variable	
C	Both A and B	

Correct Answer: A. The dependent variable

For classification problem, the dependent variable is categorical. In case of the regression problem, the dependent variable is continuous.

Prev
Module Test (/machine-learning/module-test/329)

Next
Download Certificate (/certificate/preview/machine-learning/coc/learn?score=true)

Question 2

2 Marks

Which of the following is an ensemble method?

A Decision tree

B Random forest ✓

C K-means

D Both A and B

Correct Answer: B. Random forest

Only random forest is an ensemble technique. K-means is a clustering method.

Question 3

3 marks

The best fit line in linear regression is determined using which method?

A Least square error ✓

B Model accuracy

C Logarithmic loss

D Both A and B

Correct Answer: A. Least square error

In linear regression, we try to minimize the least square errors of the model to identify the line of best fit.

Question 4

1 Mark

Clustering is a supervised machine learning algorithm.

A True

B False

Correct Answer: B. False

Clustering is an unsupervised machine learning algorithm.

Question 5

2 Marks

For selecting the value of k in K-mean algorithm, we can use which of the following method?

A Set the number of data points as k.

B Use elbow method to find k.

C We do not have to set the k value.

Correct Answer: B. Use elbow method to find k.

Out of the given options, the elbow method is used for finding the optimal number of clusters.

Prev Module Test (machine-learning/module-test/29)

Next Download Certificate (/certificate/preview/machine-learning/coc/learn?score=true)

Question 6

3 marks

Which of the following are examples of categorical variables?

1. Income
2. Gender
3. Age
4. Marital Status
5. Birth Place

A 2 and 3

B 1 and 4

C 2, 4, 5 ✓

D 1, 2, 3, 4, 5

Correct Answer: C. 2, 4, 5

Age and income are continuous variables while others are categorical.

Question 7 ✓

1 Mark

Can the mean squared error have negative values?

A Yes

B No ✓

Correct Answer: B. No

The mean squared error cannot have negative values because we use the squares of differences.

Prev
Module Test (machine-learning/module-test/329)

Next
Download Certificate (/certificate/preview/machine-learning/coc/learn?score=true)

Question 8 ✓

2 Marks

Scaling can only be applied to?

A Numerical variables ✓

B Categorical variables

C Date-time variables

D Both A and B

Correct Answer: A. Numerical variables

Scaling cannot be used with categorical variables or date-time variables.

Question 9 ✓

3 marks

Which of the following can be used to deal with categorical variables?

A One hot encoding

B Label encoding

C Log transformation

D Both A and B ✓

Correct Answer: D. Both A and B

One hot and label encoding are both used for categorical variables. Log transformation is used to transform continuous variables.

Question 10

1 Mark

Clustering is an _____ algorithm.

Final Test

H

A supervised

B unsupervised

Correct Answer: B. unsupervised

Clustering is a technique that is used to deal with unsupervised ML problems.

Question 11

2 Marks

Random forest is built using?

A Multiple linear regression

B Multiple kNN models

C Multiple decision tree

Correct Answer: C. Multiple decision tree

Random forest is an ensemble of multiple decision trees.

Question 12

3 marks

Standard deviation is robust to outliers.

A True

B False

Prev
Module Test (machine-learning/module-test/329)

Next
Download Certificate (/certificate/preview/machine-learning/coc/learn?score=true)

Correct Answer: B. False

Standard deviation is calculated considering all the points in the variable. If we have outliers in the dataset, the mean will be affected and in turn, the standard deviation will also be affected by the presence of outliers. Hence, standard deviation is not robust to outliers.

Question 13

1 Mark

Which of the following evaluation metrics can be used for classification problems?

1. RMSE
2. Accuracy
3. Log loss
4. MAE
5. Confusion Matrix

A 1, 4, 5

B 2, 3, 5

C 1, 4, 3

D 1, 2, 3

Correct Answer: B. 2, 3, 5

For classification problems, error related evaluation metrics such as RMSE, MAE, and r-squared cannot be used.

Question 14

2 Marks

Which of the following approach can be used to detect outliers in a particular variable?

A Scatter plot

B Box plot ✓

C Frequency table

Correct Answer: B. Box plot

Box plot uses whiskers to show the range of the data. Points past the end of the whiskers are the outlier points in that variable.

Question 15 ✓

3 marks

Clustering cannot be used when there is only 1 feature.

A True

B False ✓

Correct Answer: B. False

Minimum number of variables required to perform clustering analysis is 1.

Question 16 ✓

1 Mark

Which of the following evaluation metrics can be used for regression problems?

1. RMSE
2. Accuracy
3. Log loss
4. MAE
5. R-squared

A 1, 4, 3

B 2, 3, 4

C 1, 4, 5

D 1, 2, 3

Correct Answer: C. 1, 4, 5

Root mean squared error (RMSE), mean absolute error (MAE), and r-squared are used for regression problems. While accuracy, log loss, and confusion metrics can be used only when the target variable is continuous.

Question 17

2 Marks

Which of the below mentioned algorithms can be used for both classification and regression tasks?

1. Linear regression
2. Logistic regression
3. kNN
4. Decision tree
5. Random forest

A 4 and 5

B 3, 4, 5

C 1, 2, 3, 4, 5

D 2, 4, 5

Correct Answer: B. 3, 4, 5

Linear regression can only be used for regression problems while the logistic regression can only be used for classification problems. All the other techniques can be used with both regression and classification tasks.

Question 18

3 marks

Which of the following is true for r-squared and adjusted r-squared?

A Both r-squared and adjusted r-squared always increase with the addition of new features.

B R-squared may decrease on the addition of new features.

C Adjusted r-squared may decrease on addition of new features. ✓

D Both r-squared and adjusted r-squared always decrease with the addition of new features.

Correct Answer: C. Adjusted r-squared may decrease on addition of new features.

R-squared can never decrease while the adjusted r-squared may increase or decrease on the addition of new features.

Question 19

1 Mark

Which of the following is an unsupervised machine learning algorithm?

A kNN

B K-means ✓

C Decision tree

D Both A and B

Correct Answer: B. K-means

Out of the given options, only k-means is an unsupervised machine learning algorithm.

Question 20

2 Marks

Which of the following data structure(s) can be included in a dictionary?

A Tuples

B Dictionary

C Lists

D All of the above

Correct Answer: D. All of the above

Dictionaries can include tuples, dictionaries as well as lists.

Question 21

3 marks

As the value for k in kNN increases, the model starts to?

A Overfit

B Underfit

C Cannot be determined

Correct Answer: B. Underfit

The model performs poorly on both training and test data.

Prev
Module Test (machine-learning/module-test/329)

Next
Download Certificate (/certificate/preview/machine-learning/coc/learn?score=true)

Question 22

1 Mark

If we write 3*5 in the python console, we will get?

A 15

B 8

C 2

D Error

Correct Answer: A. 15

* is a multiplication operator and hence the output is 15.

Question 23

2 Marks

The AUC-ROC can be used for both classification as well as regression problems.

A True

B False

Correct Answer: B. False

The AUC ROC can only be used in case of binary classification problems.

Question 24

3 marks

Which of the following methods is used to avoid overfitting in a decision tree?

A Reducing the max depth value of the tree.

B Increasing the max depth value for the tree.

C Reducing the number of training points.

D Increasing the number of training points.

Correct Answer: A. Reducing the max depth value of the tree.

Increasing the growth of the tree increases the chances of overfitting and hence if we stop the growth, it can help us to avoid overfitting.

Question 25

1 Mark

The value of k in kNN algorithm can be decided using which of the following methods?

A By keeping k same as the number of data points.

B By plotting the elbow curve.

C The value of k will be equal to the number of features.

D The value of k in kNN cannot be changed.

Correct Answer: B. By plotting the elbow curve.

The value of k is decided using the elbow curve method.

Question 26

2 Marks

Can logistic regression be used for multiclass classification problems?

A Yes

B No

Correct Answer: A. Yes

The logistic regression can be implemented for a multiclass problem using the one-versus all method.

Question 27

3 marks

Which of the following is not a function of dataframe?

A dataframe.columns

B dataframe.rows

C dataframe.tail()

D dataframe.shape

Correct Answer: B. dataframe.rows

.rows is not a function of dataframe whereas all the remaining options represent a function of a dataframe.

Question 28

1 Mark

The best split in a decision tree algorithm is decided based on the purity of the node.

A True

B False

Prev
Module Test (machine-learning/module-test/329)

Next
Download Certificate (/certificate/preview/machine-learning/coc/learn?score=true)

Correct Answer: A. True

The objective of a decision tree algorithm is to create the most pure nodes and classify all the data points correctly.

Question 29

2 Marks

Which of the following can be used to scale the variables?

A Min max scaler

B Standard scaler

C Both A and B

D None of the above

Correct Answer: C. Both A and B

The data can be scaled using either of the two methods.

Question 30

3 marks

Only categorical variables can be used for splitting in case of a decision tree algorithm.

A True

B False

Correct Answer: B. False

Both categorical and continuous variables are considered while deciding the best split in the decision tree.

Question 31

1 Mark

Linear regression and logistic regression are both used for regression problems.

A True

B False

Correct Answer: B. False

Linear regression is used for regression problems while the logistic regression can only be used for classification problems.

Question 32

2 Marks

Random forest can handle missing values on its own.

A Yes

B No

Correct Answer: B. No

We need to impute the missing values before we build a random forest model.

Question 33

3 marks

Which of the following algorithm is affected by the scale of data?

A KNN

B K-means

C Decision tree

D Both A and B ✓

Correct Answer: D. Both A and B

Decision tree is a tree based algorithm while the other two are distance based algorithm.

Question 34 ✓

1 Mark

Which of the following can be used to check the shape of the dataset (given that the dataset is stored with the name df)?

A df.dtypes

B df.shape()

C df.shape ✓

D None of the above

Correct Answer: C. df.shape

To check the shape of the dataset, we use the dataset_name.shape. Thus option C is correct.

Question 35 ✓

2 Marks

What will be the output of the below statement?

```
if 5 > 10
print("fan")

elif 8 != 9:

print("glass")

else:
print("cream")
```

A fan

B glass ✓

C cream

D 5

Correct Answer: B. glass

Out of all the conditions, 8 != 9 (8 is not equal to 9) is true. So, elif statement will be true and thus the output will be ‘glass’.