

Н

Score: 94%

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

Download Certificate

Final Test Summary

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

Show incorrect attempt only

1 Mark Question 1

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

Α The dependent variable

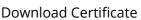
В The independent variable

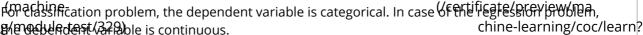
C Both A and B

Gorpert Answer: A. The dependent variable

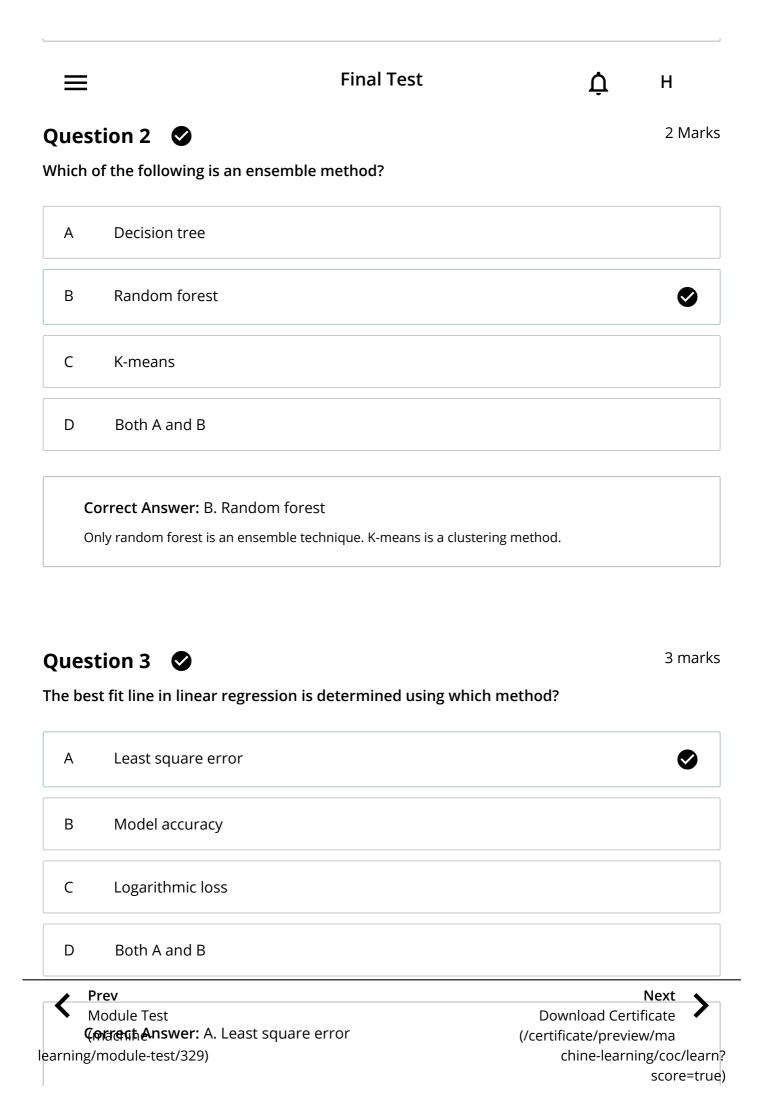
learning/engelydectest/3229ble is continuous.

Next





score=true)



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.

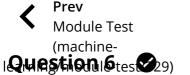
B Use elbow method to find k.

lacksquare

C We do not have to set the k value.

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

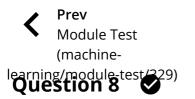
Out of the given options, the elbow method is used for finding the optimal number of clusters.

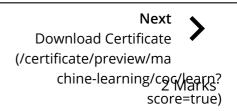


Next
Download Certificate

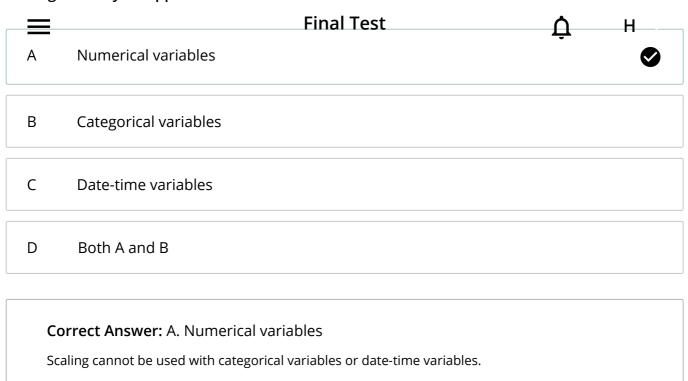
(/certificate/preview/ma
chine-learning/coc/learn?
score=true)

Which of the following are example 1. Income 2. Gender 3. Age 4. Marital Status 5. Birth Place		Final Test	Ţ	Н
А	2 and 3			
В	1 and 4			
С	2, 4, 5			Ø
D	1, 2, 3, 4, 5			
	orrect Answer: C. 2, 4, 5	ous variables while others are categorical.		
	tion 7 🔗 e mean squared error h	ave negative values?		1 Mark
Α	Yes			
В	No			•
	orrect Answer: B. No	not have negative values because we use th	e squares of differen	nces





Scaling can only be applied to?

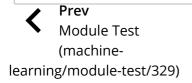


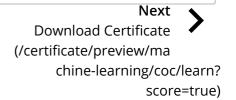
Which of the following can be used to deal with categorical variables?

- 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.





B Multiple kNN models

C Multiple decision tree

lacktriangle

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.



Prev Modulle est (machineNext
Download Certificate
(/certificate/preview/ma-





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



Correct Answer: B. 2, 3, 5

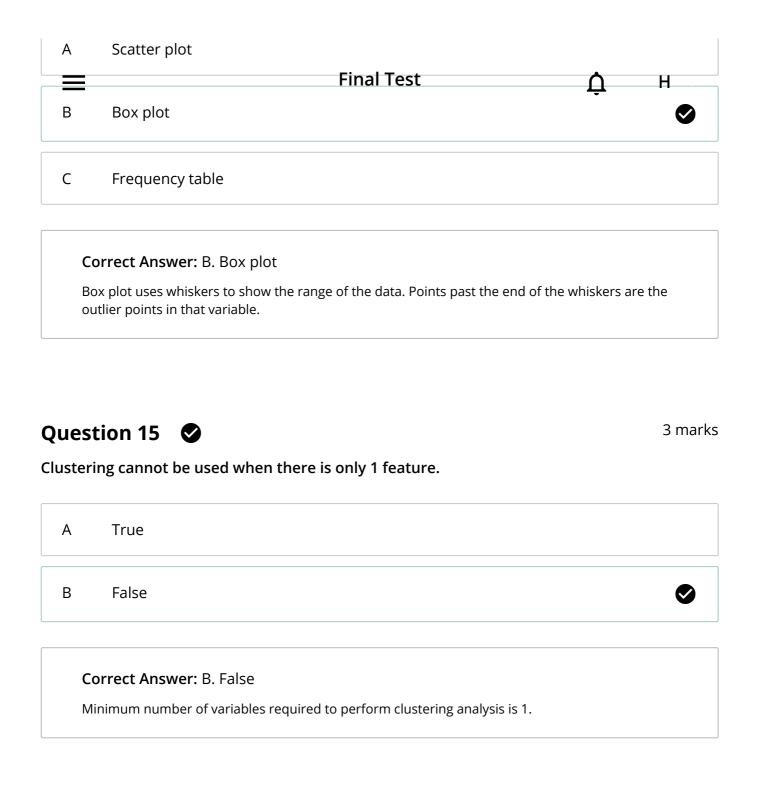
For classification problems, error related evaluation metrics such as RMSE, MAE, and r-squared cannot be used.

Next Marks

Download Certificate

Which of the following approach can be used to detect outliers in a particular variable? learning/module-test/329)

chine-learning/coc/learn?



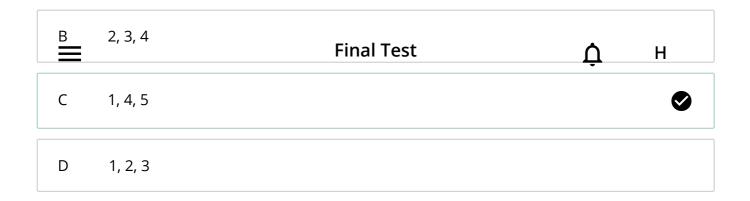
Question 16

1 Mark

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

- 1. RMSE
- 2. Accuracy
- 3. Log loss
- 4. MAE
- 5. B-squared
 Module Test
 Download Certificate

 (machine(machinelear#ing/module3test/329)
 chine-learning/coc/learn?
 score=true)



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.

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



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

Н

- Α Both r-squared and adjusted r-squared always increase with the addition of new features.
- R-squared may decrease on the addition of new features. В
- C Adjusted r-squared may decrease on addition of new features.



Both r-squared and adjusted r-squared always decrease with the addition of D 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?

Α kNN

K-means В



C Decision tree

Both A and B D

learning/module-test/329)

Next

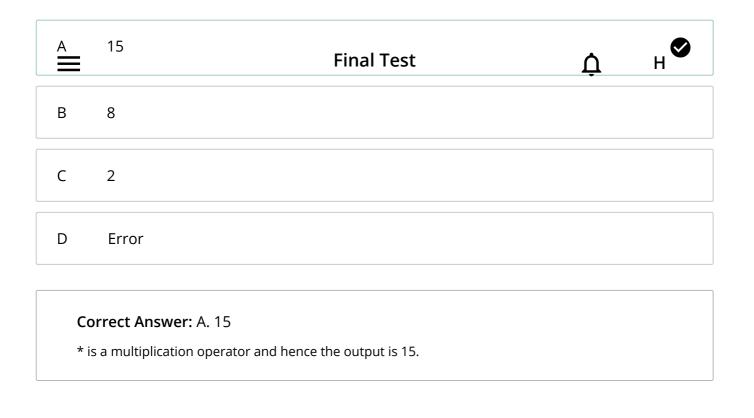
Download Certificate





The model performs poorly on both training and test data.

Correct Answer: B. Underfit



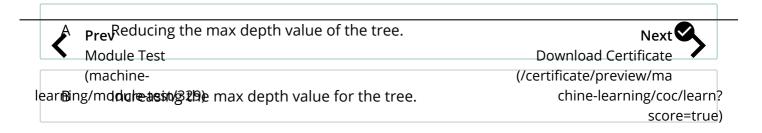
The AUC-ROC can be used for both classification as well as regression problems.



Correct Answer: B. False

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

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



<u>C</u>	Reducing the number of training points. Final Test	Û	Н	
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.

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.



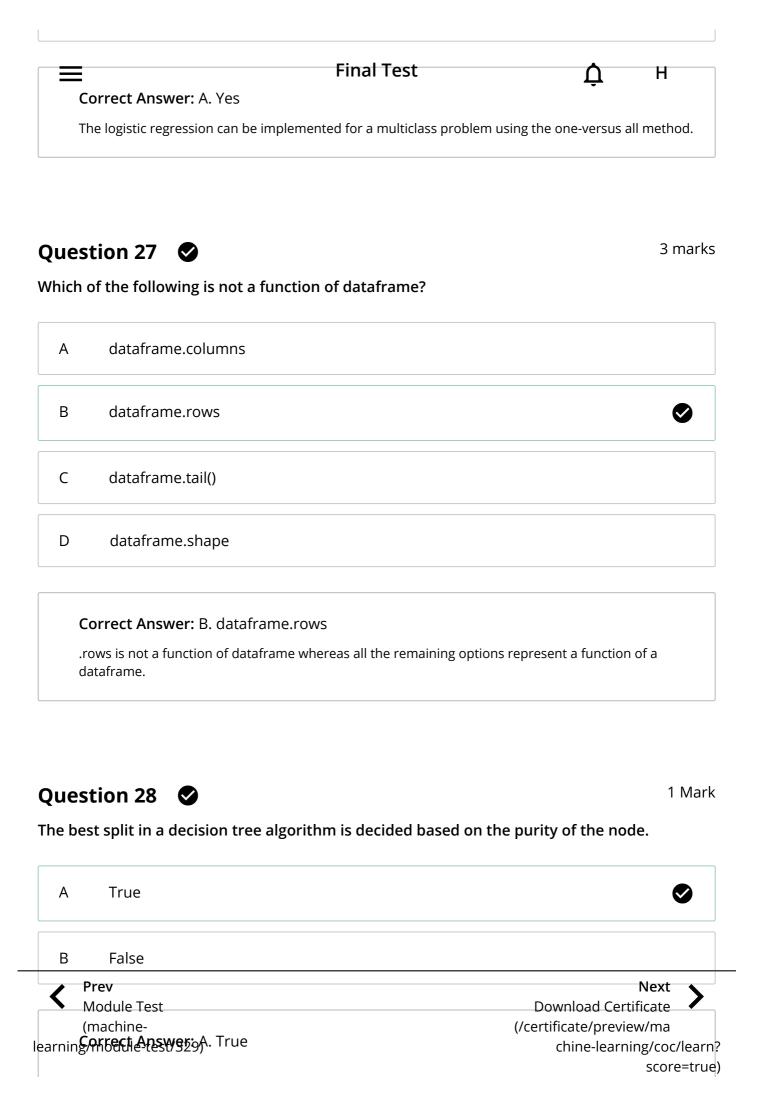
- C The value of k will be equal to the number of features.
- D The value of k in kNN cannot be changed.

Correct Answer: B. By plotting the elbow curve.

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

Can logistic regression be used for multiclass classification problems?





Question 30 amarks

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

The decision

Download Certificate

(/certificate/preview/ma

chine-learning/coc/learn?

score=true)

learning/module-test/329)



Correct Answer: D. Both A and B

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

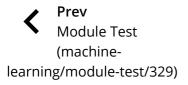
Which of the following can be used to check the shape of the dataset (given that the dataset is stored with the name df)?

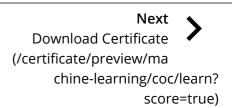


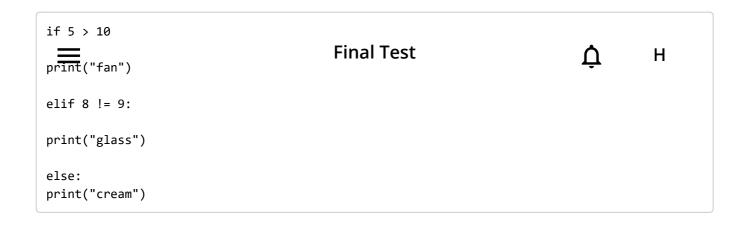
Correct Answer: C. df.shape

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

What will be the output of the below statement?







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'.