E20-007 Q&As

Data Science and Big Data Analytics

Pass EMC E20-007 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass4itsure.com/e20-007.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by EMC Official Exam Center

QUESTION 1

In linear regression modeling, which action can be taken to improve the linearity of the relationship between the dependent and independent variables?

A. Apply a transformation to a variable

B. Use a different statistical package

C. Calculate the R-Squared value

D. Change the units of measurement on the independent variable

Correct Answer: A

QUESTION 2

A data scientist plans to classify the sentiment polarity of 10, 000 product reviews collected from the Internet. What is the most appropriate model to use? Suppose labeled training data is available.

A. Na?e Bayesian classifier

B. Linear regression

C. Logistic regression

D. K-means clustering

Correct Answer: A

QUESTION 3

In a Student\\'s t-test, what is the meaning of the p-value?

A. Area under the appropriate tails of the student\\'s distribution

B. Power of the Student\\'s t-test

C. Mean of the distribution for the null hypothesis

D. Mean of the distribution for the alternate hypothesis

Correct Answer: A

QUESTION 4

You have fit a decision tree classifier using 12 input variables. The resulting tree used 7 of the 12 variables, and is 5 levels deep. Some of the nodes contain only 3 data points. The AUC of the model is

0.85. What is your evaluation of this model?

A. The tree is probably overfit. Try fitting shallower trees and using an ensemble method.

B. The AUC is high, and the small nodes are all very pure. This is an accurate model.

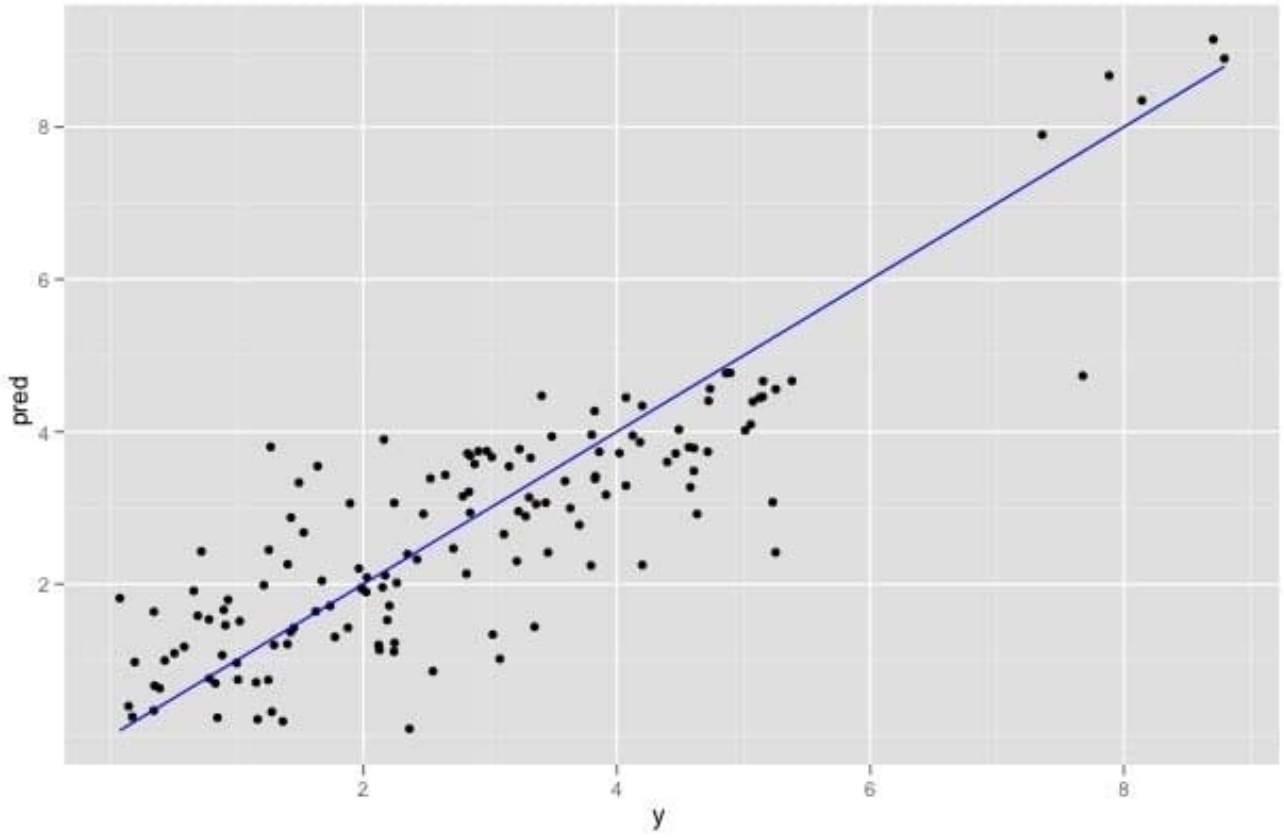
C. The tree did not split on all the input variables. You need a larger data set to get a more accurate model.

D. The AUC is high, so the overall model is accurate. It is not well-calibrated, because the small nodes will give poor estimates of probability.

Correct Answer: A

QUESTION 5

Refer to the exhibit.



You have run a linear regression model against your data, and have plotted true outcome versus predicted outcome. The R-squared of your model is 0.75. What is your assessment of the model?

A. The R-squared may be biased upwards by the extreme-valued outcomes. Remove them and refit to get a better idea of the model\\'s quality over typical data.

B. The R-squared is good. The model should perform well.

C. The extreme-valued outliers may negatively affect the model\\'s performance. Remove them to see if the R-squared

improves over typical data.

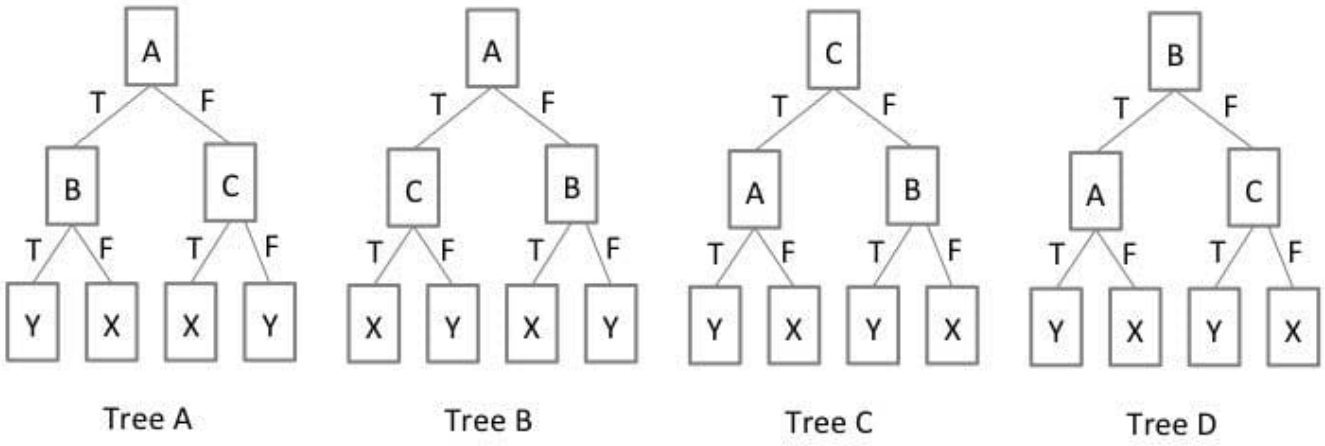
D. The observations seem to come from two different populations, but this model fits them both equally well.

Correct Answer: A

QUESTION 6

Refer to the Exhibit.

A	B	C	CLASS
T	T	T	X
T	T	F	Y
T	F	T	X
F	F	F	Y
F	T	T	X
F	F	T	Y



In the Exhibit, the table shows the values for the input Boolean attributes "A", "B", and "C". It also shows the values for the output attribute "class". Which decision tree is valid for the data?

A. Tree B

B. Tree A

C. Tree C

D. Tree D

Correct Answer: A

QUESTION 7

The average purchase size from your online sales site is $17, 200. The customer experience team believes a certain adjustment of the website will increase sales. A pilot study on a few hundred customers showed an increase in average purchase size of $1.47, with a significance level of p=0.1.

The team runs a larger study, of a few thousand customers. The second study shows an increased average purchase size of $0.74, with a significance level of 0.03. What is your assessment of this study?

A. The change in purchase size is not practically important, and the good p-value of the second study is probably a result of the large study size.

B. The change in purchase size is small, but may aggregate up to a large increase in profits over the entire customer base.

C. The difference in the change in purchase size between the two studies is troubling; The team should run another, larger study.

D. The p-value of the second study shows a statistically significant change in purchase size. The new website is an improvement.

Correct Answer: A

QUESTION 8

Refer to the exhibit.

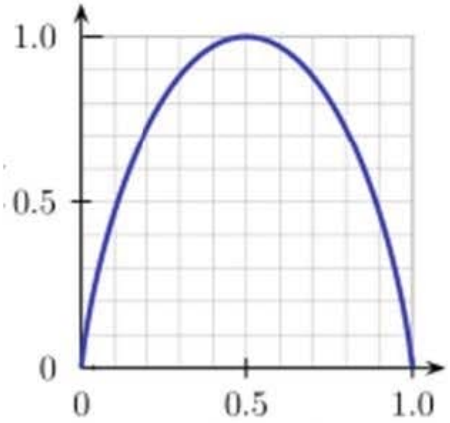


Fig-A

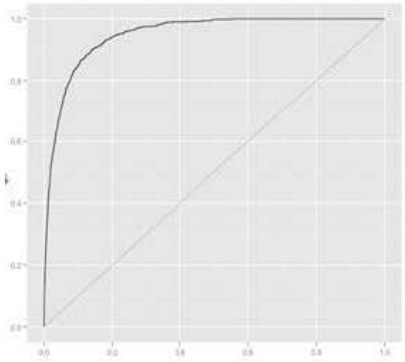


Fig-B

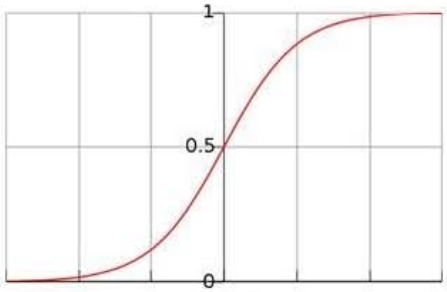


Fig-C

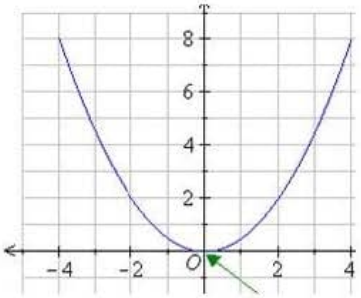


Fig-D

$$H = -\sum_{c} p(c) \log_2 p(c)$$

p (c) is the probability of a given class

The exhibit shows four graphs labeled as Fig A thorough Fig D. Which figure represents the entropy function relative to a Boolean classification and is represented by the formula shown in Exhibit?

A. Fig-A

B. Fig-B

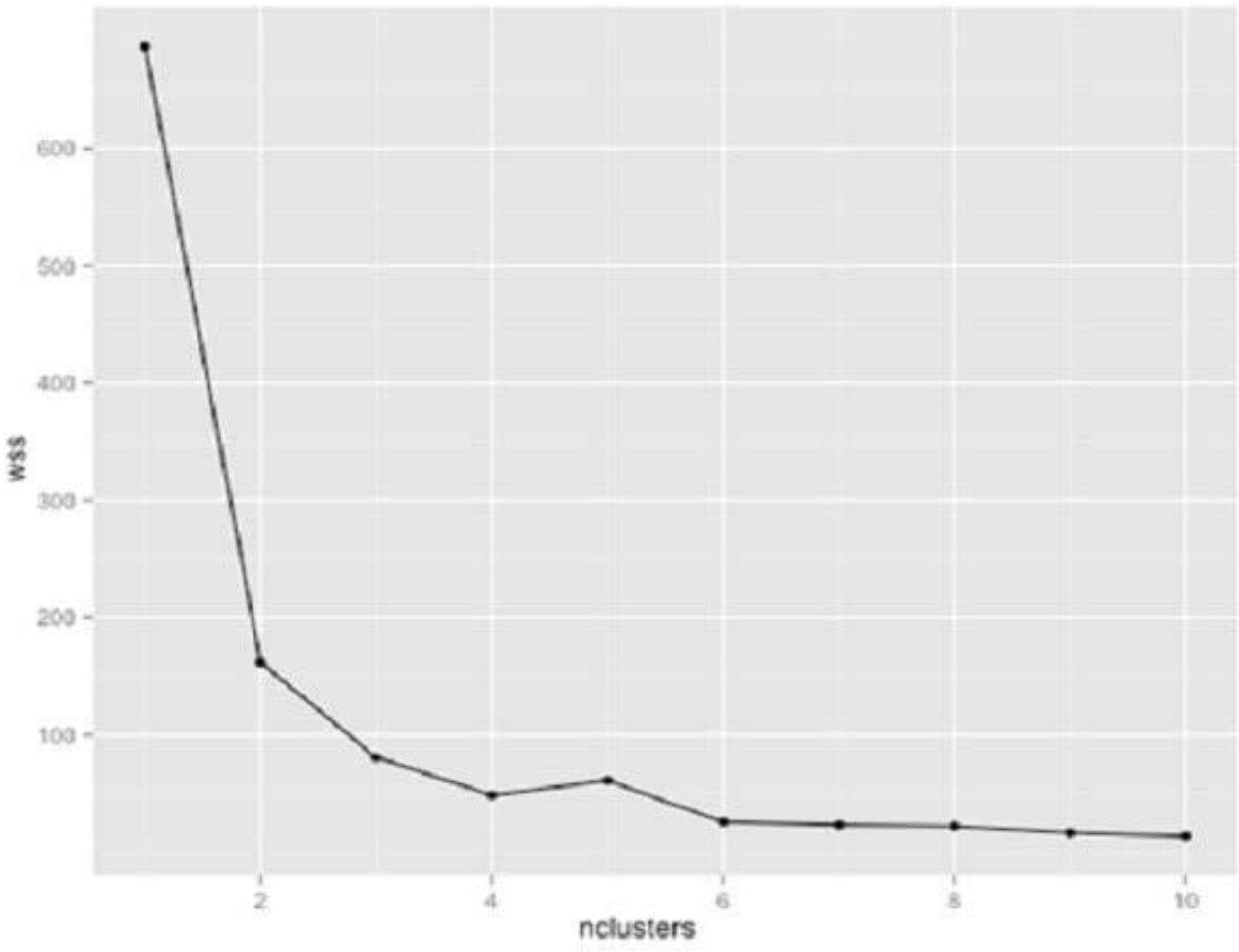
C. Fig-C

D. Fig-D

Correct Answer: A

QUESTION 9

Refer to the exhibit.



You are using K-means clustering to classify customer behavior for a large retailer. You need to determine the optimum number of customer groups. You plot the within-sum-of- squares (wss) data as shown in the exhibit. How many customer groups should you specify?

A. 2

B. 3

C. 4

D. 8

Correct Answer: C

QUESTION 10

Which method is used to solve for coefficients b0, b1, .., bn in your linear regression model : Y = b0 + b1x1+b2x2+....+bnxn

A. Ordinary Least squares

B. Apriori Algorithm

C. Ridge and Lasso

D. Integer programming

Correct Answer: A

QUESTION 11

Refer to the exhibit Consider the training data set shown in the exhibit. What are the classification (Y = 0 or 1) and the probability of the classification for the tupleX(0, 0, 1) using Naive Bayesian classifier?

X1	X2	X3	Y
1	1	1	0
1	1	0	0
0	0	0	0
0	1	0	1
1	0	1	1
0	1	1	1

A. Classification Y = 1, Probability = 4/54

B. Classification Y = 0, Probability = 1/54

C. Classification Y = 1, Probability = 1/54

D. Classification Y = 0, Probability = 4/54

Correct Answer: A

QUESTION 12

Which word or phrase completes the statement; "A data scientist would consider a RDBMS is to a table as R is to a _____."?

A. Data frame

B. List

C. Matrix

D. Array

Correct Answer: A

QUESTION 13

Your company has 3 different sales teams. Each team\\'s sales manager has developed incentive offers to increase the size of each sales transaction. Any sales manager whose incentive program can be shown to increase the size of the average sales transaction will receive a bonus.

Data are available for the number and average sale amount for transactions offering one of the incentives as well as transactions offering no incentive.

The VP of Sales has asked you to determine analytically if any of the incentive programs has resulted in a demonstrable increase in the average sale amount. Which analytical technique would be appropriate in this situation?

A. One-way ANOVA

B. Multi-way ANOVA

C. Student\\'s t-test

D. Wilcoxson Rank Sum Test

Correct Answer: A

QUESTION 14

What does the R code

z