



DATABRICKS-CERTIFIED- PR OFESIONAL-DATA-SCIENTIST^{Q&As}

Databricks Certified Professional Data Scientist Exam

**Pass Databricks DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST Exam with 100%
Guarantee**

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

<https://www.geekcert.com/databricks-certified-professional-data-scientist.html>

100% Passing Guarantee
100% Money Back Assurance

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



VCE & PDF

GeekCert.com

<https://www.geekcert.com/databricks-certified-professional-data-scientist.html>
2024 Latest geekcert DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF and VCE dumps Download

- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers





QUESTION 1

Which of the following is a correct example of the target variable in regression (supervised learning)?

- A. Nominal values like true, false
- B. Reptile, fish, mammal, amphibian, plant, fungi
- C. Infinite number of numeric values, such as 0.100, 42.001, 1000.743..
- D. All of the above

Correct Answer: D

Explanation: We address two cases of the target variable. The first case occurs when the target variable can take only nominal values: true or false; reptile, fish: mammal, amphibian, plant, fungi. The second case of classification occurs when the target variable can take an infinite number of numeric values, such as 0.100, 42.001, 1000.743, This case is called regression.

QUESTION 2

Refer to exhibit

Independent Variable	Coefficient	P-Value
A	0.45	0
B	3.67	0
C	1.23	0

$$R^2 = 0.10$$

You are asked to write a report on how specific variables impact your client's sales using a data set provided to you by the client. The data includes 15 variables that the client views as directly related to sales, and you are restricted to these variables only. After a preliminary analysis of the data, the following findings were made: 1. Multicollinearity is not an issue among the variables 2. Only three variables-A, B, and C-have significant correlation with sales You build a linear regression model on the dependent variable of sales with the independent variables of A, B, and C. The results of the regression are seen in the exhibit. You cannot request additional data. what is a way that you could try to increase the R2 of the model without artificially inflating it?

- A. Create clusters based on the data and use them as model inputs
- B. Force all 15 variables into the model as independent variables
- C. Create interaction variables based only on variables A, B, and C
- D. Break variables A, B, and C into their own univariate models

Correct Answer: A



Explanation: In statistics, linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variable) denoted X . The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. (This term should be distinguished from multivariate linear regression[^] where multiple correlated dependent variables are predicted, rather than a single scalar variable.) In linear regression data are modeled using linear predictor functions, and unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, linear regression refers to a model in which the conditional mean of y given the value of X is an affine function of X . Less commonly: linear regression could refer to a model in which the median, or some other quantile of the conditional distribution of y given X is expressed as a linear function of X . Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of y given X , rather than on the joint probability distribution of y and X : which is the domain of multivariate analysis.

QUESTION 3

A denote the event '\student is female\' and let B denote the event '\student is French\''. In a class of 100 students suppose 60 are French, and suppose that 10 of the French students are females. Find the probability that if I pick a French student, it will be a girl, that is, find $P(A|B)$.

- A. $1/3$
- B. $2/3$
- C. $1/6$
- D. $2/6$

Correct Answer: C

Explanation: Since 10 out of 100 students are both French and female, then $P(A \text{ and } B) = 10/100$ Also. 60 out of the 100 students are French, so $P(B) = 60/100$ So the required probability is: $P(A|B) = P(A \text{ and } B) / P(B) = 10/100 \div 60/100 = 1/6$

QUESTION 4

Which of the following technique can be used to the design of recommender systems?

- A. Naive Bayes classifier
- B. Power iteration
- C. Collaborative filtering
- D. 1 and 3
- E. 2 and 3

Correct Answer: C

Explanation: One approach to the design of recommender systems that has seen wide use is collaborative filtering. Collaborative filtering methods are based on collecting and analyzing a large amount of information on users\' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender systems. For



example the k-nearest neighbor (k-NN) approach and the Pearson Correlation

QUESTION 5

The method based on principal component analysis (PCA) evaluates the features according to:

- A. The projection of the largest eigenvector of the correlation matrix on the initial dimensions
- B. According to the magnitude of the components of the discriminate vector
- C. The projection of the smallest eigenvector of the correlation matrix on the initial dimensions
- D. None of the above

Correct Answer: A

Explanation: Feature Selection:

The method based on principal component analysis (PCA) evaluates the features according to the projection of the largest eigenvector of the correlation matrix on the initial dimensions, the method based on Fisher's linear discriminate analysis evaluates. Then according to the magnitude of the components of the discriminate vector.

[Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps](#)

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF Dumps](#)

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Braindumps](#)