DATABRICKS-CERTIFIED-PROFESSIONAL-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

QUESTION 1

What describes a true property of Logistic Regression method?

A. It handles missing values well.

B. It works well with discrete variables that have many distinct values.

C. It is robust with redundant variables and correlated variables.

D. It works well with variables that affect the outcome in a discontinuous way.

Correct Answer: C

QUESTION 2

You are working with the Clustering solution of the customer datasets. There are almost 40 variables are available for each customer and almost 1.00,0000 customer\\'s data is available. You want to reduce the number of variables for clustering, what would you do?

A. You will randomly reduce the number of variables

B. You will find the correlation among the variables and from their variables are not co- related will be discarded.

C. You will find the correlation among the variables and from the highly co-related variables, you will be considering only one or two variables from it.

D. You cannot discard any variable for creating clusters.

E. You can combine several variables in one variable

Correct Answer: CE

Explanation: When you are applying clustering technique and you find that there are quite a huge number of variables are available. Then it is better the find the co-relation among the variables and consider only one or two variables from the highly co-related variables. Because highly co-related variable will have the same effect, while creating the cluster. We can use scatter plot matrix among the variables to find the co-relation. You can also combine several variables into a single variable. For example if you have two values in the dataset like Asset and Debt than by combining these two values like Debt to Asset ratio and use it while creating the cluster.

QUESTION 3

In which of the scenario you can use the linear regression model?

A. Predicting Home Price based on the location and house area

B. Predicting demand of the goods and services based on the weather

C. Predicting tumor size reduction based on input as number of radiation treatment

D. Predicting sales of the text book based on the number of students in state

Correct Answer: ABCD

Explanation: : You can use the linear regression model for predicting the continuous output variable based on the input variables. In all the cases mentioned in the question option, you can see that output can be predicted based on the input variable. Option-A: Input: Location, House Area and Output: House Price Option-B : Input: Weather condition, Output: Demand for the goods and services Option-C : Input: Number of Radiation Session Output: Tumor Size Reduction Option-D : Input: Number of students and Output: Sale quantity of text book

QUESTION 4

You are having 1000 patients\\' data with the height and age. Where age in years and height in meters. You wanted to create cluster using this two attributes. You wanted to have near equal effect for both the age and height while creating the cluster. What you can do?

A. You will be adding height with the numeric value 100

B. You will be converting each height value to centimeters

C. You will be dividing both age and height with their respective standard deviation

D. You will be taking square root of height

Correct Answer: BC

Explanation: When you see the data age in years would have values like 50, 60r 70 90 years etc. And while calculating distance from centroid maximum possible value can be 90- 0 and its square will be 8100.

While using heights in meter can be 2-0.5(1.5) meters and its square will be 2.25 only. So you can see age has more effect than height. Hence bringing the height on same level you can convert it into centimeters. Can bring data upto 200

centimeters and then it be more effective like square of 200 maximum.

However there is another approach is to divide the each value with its standard deviation, which will not have impact of the units e.g. age/sd of the age, which results in value without unit. This can also help in reducing the effect of units.

QUESTION 5

In which phase of the data analytics lifecycle do Data Scientists spend the most time in a project?

A. Discovery

B. Data Preparation

C. Model Building

D. Communicate Results

Correct Answer: B