

DATABRICKS-CERTIFIED-PR OFESSIONAL-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.pass4itsure.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 https://www.pass4itsure.com/databricks-certified-professional-data-scientis 2024 Latest pass4itsure DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF and VCE dumps Download

- Instant Download After Purchase
- 100% Money Back Guarantee
- 😳 365 Days Free Update

VCE & PDF

Pass4itSure.com

800,000+ Satisfied Customers





QUESTION 1

Which of the following are advantages of the Support Vector machines?

- A. Effective in high dimensional spaces.
- B. it is memory efficient
- C. possible to specify custom kernels
- D. Effective in cases where number of dimensions is greater than the number of samples
- E. Number of features is much greater than the number of samples, the method still give good performances
- F. SVMs directly provide probability estimates

Correct Answer: ABCD

Explanation: Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and outliers detection.

The advantages of support vector machines are:

Effective in high dimensional spaces.

Still effective in cases where number of dimensions is greater than the number of samples.

Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels. The disadvantages of support vector machines include:

If the number of features is much greater than the number of samples, the method is likely to give poor performances.

SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation.

QUESTION 2

Which technique you would be using to solve the below problem statement? "What is the probability that individual customer will not repay the loan amount?"

- A. Classification
- B. Clustering
- C. Linear Regression
- **D.** Logistic Regression
- E. Hypothesis testing

Correct Answer: D



QUESTION 3

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 4

Which of the following statement is true for the R square value in the regression model?

- A. When R square =1, all the residuals are equal to 0
- B. When R square =0, all the residual are equal to 1
- C. R square can be increased by adding more variables to the model.
- D. R-squared never decreases upon adding more independent variables.

Correct Answer: ACD

Explanation: R square can be made high, it means when we add more variables R-square will increase. And R-square will never decreases if you add more independent variables. Higher R square value can have lower the residuals.

QUESTION 5

You are doing advanced analytics for the one of the medical application using the regression and you have two variables which are weight and height and they are very important input variables, which cannot be ignored and they are also

highly co-related.



What is the best solution for that?

- A. You will take cube root of height
- B. You will take square root of weight
- C. You will take square of the height.
- D. You would consider using BMI (Body Mass Index)

Correct Answer: D

Explanation: If multiple variables are highly co-related then it is better you consider using the either of the variable which correlates more (which is not in the given option) or go for the new variable which is a function of the both the variable in this case it could be BMI (Body Mass Index). Because it is a function of both weight and height as per the below formula. BMI = Weight/(Height * Height)

DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST PDF DumpsDATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST Exam
QuestionsDATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST Braindumps