

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 method is used to solve for coefficients bO, b1, ... bn in your linear regression model:



- A. Apriori Algorithm
- B. Ridge and Lasso
- C. Ordinary Least squares
- D. Integer programming

Correct Answer: C

Explanation: : RY = b0 + b1x1+b2x2+ + bnxn In the linear model, the bi\\'s represent the unknown p parameters. The estimates for these unknown parameters are chosen so that, on average, the model provides a reasonable estimate of a person\\'s income based on age and education. In other words, the fitted model should minimize the overall error between the linear model and the actual observations. Ordinary Least Squares (OLS) is a common technique to estimate the parameters

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

Refer to the exhibit.

Attribute	Info-Gain
Age	0.0310
Income	0.0100
Gender	0.0034
Credit Score	0.0456

You are building a decision tree. In this exhibit, four variables are listed with their respective values of info-gain. Based on this information, on which attribute would you expect the next split to be in the decision tree?

- A. Credit Score
- B. Age
- C. Income
- D. Gender

Correct Answer: A

QUESTION 4

Your customer provided you with 2. 000 unlabeled records three groups. What is the correct analytical method to use?

A. Semi Linear Regression

- B. Logistic regression
- C. Naive Bayesian classification
- D. Linear regression
- E. K-means clustering

Correct Answer: E

Explanation: k-means clustering is a method of vector quantization^ originally from signal processing, that is popular for cluster analysis in data mining, k-means clustering aims to partition n observations into k clusters in which each

observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster This results in a partitioning of the data space into Voronoi cells.

The problem is computationally difficult (NP-hard); however there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for

mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally they both use cluster centers to model the data; however k-means clustering tends to find clusters of comparable spatial extent,



while the expectation-maximization mechanism allows clusters to have different shapes.

The algorithm has nothing to do with and should not be confused with k-nearest neighbor another popular machine learning technique.

QUESTION 5

Spam filtering of the emails is an example of

- A. Supervised learning
- B. Unsupervised learning
- C. Clustering
- D. 1 and 3 are correct
- E. 2 and 3 are correct

Correct Answer: A

Explanation: Clustering is an example of unsupervised learning. The clustering algorithm finds groups within the data without being told what to look for upfront. This contrasts with classification, an example of supervised machine learning, which is the process of determining to which class an observation belongs. A common application of classification is spam filtering. With spam filtering we use labeled data to train the classifier: e-mails marked as spam or ham.

DATABRICKS-CERTIFIED- DATABRICKS-CERTIFIED- DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF Dumps

PROFESSIONAL-DATA-**SCIENTIST Practice Test**

PROFESSIONAL-DATA-SCIENTIST Study Guide