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.leads4pass.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

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST PDF Dumps |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST VCE Dumps

1 / 5

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

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST PDF Dumps |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST VCE Dumps

2 / 5

QUESTION 1

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.

QUESTION 2

Which is an example of supervised learning?

A. PCA

B. k-means clustering

C. SVD

D. EM

E. SVM

Correct Answer: E

Explanation: SVMs can be used to solve various real world problems:

SVMs are helpful in text and hypertext categorization as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.

Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback.

SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly.

Hand-written characters can be recognized using SVM

QUESTION 3

A bio-scientist is working on the analysis of the cancer cells. To identify whether the cell is cancerous or not, there has been hundreds of tests are done with small variations to say yes to the problem. Given the test result for a sample of healthy and cancerous cells, which of the following technique you will use to determine whether a cell is healthy?

A. Linear regression

B. Collaborative filtering

C. Naive Bayes

D. Identification Test

Correct Answer: C

Explanation: In this problem you have been given high-dimensional independent variables like yes, no: test results etc. and you have to predict either valid or not valid (One of two). So all of the below technique can be applied to this problem. Support vector machines Naive Bayes Logistic regression Random decision forests

QUESTION 4

If E1 and E2 are two events, how do you represent the conditional probability given that E2 occurs given that E1 has occurred?

A. $P(E1)/P(E2)$

B. $P(E1+E2)/P(E1)$

C. $P(E2)/P(E1)$

D. $P(E2)/(P(E1+E2)$

Correct Answer: C

QUESTION 5

You have collected the 100\\'s of parameters about the 1000\\'s of websites e.g. daily hits, average time on the websites, number of unique visitors, number of returning visitors etc. Now you have find the most important parameters which can best describe a website, so which of the following technique you will use:

A. PCA (Principal component analysis)

B. Linear Regression

C. Logistic Regression

D. Clustering

Correct Answer: A

Explanation: Principal component analysis . or PCA, is a technique for taking a dataset that is in the form of a set of tuples representing points in a high-dimensional space and finding the dimensions along which the tuples line up best. The idea is to treat the set of tuples as a matrix M and find the eigenvectors for MMT or M T M . The matrix of these eigenvectors can be thought of as a rigid rotation in a high-dimensional space. When you apply this transformation to

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF Dumps |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST VCE Dumps

4 / 5

the original data, the axis corresponding to the principal eigenvector is the one along which the points are most "spread out,11 More precisely this axis is the one along which the variance of the data is maximized. Put another way, the points can best be viewed as lying along this axis, with small deviations from this axis.

Latest DATABRICKS-CERT
IFIED-PROFESSIONAL-
DATA-SCIENTIST Dumps

DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST PDF Dumps

DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST VCE Dumps

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED- PROFESSIONAL-DATA-SCIENTIST PDF Dumps | DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST VCE Dumps

5 / 5