



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.passapply.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

PassApply.com

<https://www.passapply.com/databricks-certified-professional-data-scientist.2024-Latest-passapply-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

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. Them according to the magnitude of the components of the discriminate vector.

QUESTION 2

Which of the following true with regards to the K-Means clustering algorithm?

- A. Labels are not pre-assigned to each objects in the cluster.
- B. Labels are pre-assigned to each objects in the cluster.
- C. It classify the data based on the labels.
- D. It discovers the center of each cluster.
- E. It find each objects fall in which particular cluster

Correct Answer: ADE

Explanation: Clustering does not require any predefined labels on the object, rather it consider the attributes on the object. Hence, option-B is out. Clustering is different than classification technique.

Hence you can discard the option-C as well. It does not use the pre-defined labels, hence it is called unsupervised learning and option-A is correct. Main purpose of the Clustering technique is to determine the center of each Cluster and then

find the distance from that center. If object is near the center than it would fall in that particular cluster. Hence, finally you will have group or clusters created and get to know that objects fall in which particular cluster.

QUESTION 3

$y_1, y_2, y_3, \dots, y_{n-1}, y_n$



May have a trend component that is quadratic in nature. Which pattern of data will indicate that the trend in the time series data is quadratic in nature?

- A. Naive Bayesian classifier
- B. Decision tree
- C. Linear regression
- D. K-means clustering

Correct Answer: D

Explanation: kmeans uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible. You can control the details of the minimization using several optional input parameters to kmeans, including ones for the initial values of the cluster centroids, and for the maximum number of iterations. Clustering is primarily an exploratory technique to discover hidden structures of the data, possibly as a prelude to more focused analysis or decision processes. Some specific applications of k-means are image processing, medical and customer segmentation. Clustering is often used as a lead-in to classification. Once the clusters are identified, labels can be applied to each cluster to classify each group based on its characteristics. Marketing and sales groups use k-means to better identify customers who have similar behaviors and spending patterns.

QUESTION 4

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 5



You are analyzing data in order to build a classifier model. You discover non-linear data and discontinuities that will affect the model. Which analytical method would you recommend?

- A. Logistic Regression
- B. Decision Trees
- C. Linear Regression
- D. ARIMA

Correct Answer: B

Explanation: A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represents classification rules.

In decision analysis a decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values (or expected utility) of competing alternatives are calculated.

A decision tree consists of 3 types of nodes:

1.

Decision nodes - commonly represented by squares

2.

Chance nodes - represented by circles

3.

End nodes - represented by triangles Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. If in practice decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities.

Decision trees, influence diagrams, utility functions, and other decision analysis tools and methods are taught to undergraduate students in schools of business, health economics, and public health, and are examples of operations research or management science methods.

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

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Practice Test](#)

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