

https://www.geekcert.com/databricks-certified-professional-data-scientist.ht 2024 Latest geekcert DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF and VCE dumps Download

# DATABRICKS-CERTIFIED-PR OFESSIONAL-DATA-SCIENTIST<sup>Q&As</sup>

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 VCE & PDF GeekCert.com

https://www.geekcert.com/databricks-certified-professional-data-scientist.ht 2024 Latest geekcert 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**

Select the correct option which applies to L2 regularization

- A. Computational efficient due to having analytical solutions
- B. Non-sparse outputs
- C. No feature selection

Correct Answer: ABC

Explanation: :

The difference between their properties can be promptly summarized as follows:

L2 regularization	L1 regularization
Computational efficient due to having analytical solutions	Computational inefficient on non-sparse cases
Non-sparse outputs	Sparse outputs
No feature selection	Built-in feature selection

#### **QUESTION 2**

What is the considerable difference between L1 and L2 regularization?

- A. L1 regularization has more accuracy of the resulting model
- B. Size of the model can be much smaller in L1 regularization than that produced by L2- regularization

C. L2-regularization can be of vital importance when the application is deployed in resource-tight environments such as cell-phones.

D. All of the above are correct

#### Correct Answer: B

Explanation: The two most common regularization methods are called L1 and L2 regularization. L1 regularization penalizes the weight vector for its L1-norm (i.e. the sum of the absolute values of the weights), whereas L2 regularization uses its L2-norm. There is usually not a considerable difference between the two methods in terms of the accuracy of the resulting model (Gao et al 2007), but L1 regularization has a significant advantage in practice. Because many of the weights of the features become zero as a result of L1- regularized training, the size of the model can be much smaller than that produced by L2- regularization. Compact models require less space on memory and storage, and enable the application to start up quickly. These merits can be of vital importance when the application is deployed



in resource-tight environments such as cell-phones. Regularization works by adding the penalty associated with the coefficient values to the error of the hypothesis. This way, an accurate hypothesis with unlikely coefficients would be penalized whila a somewhat less accurate but more conservative hypothesis with low coefficients would not be penalized as much.

# **QUESTION 3**

You have used k-means clustering to classify behavior of 100, 000 customers for a retail store. You decide to use household income, age, gender and yearly purchase amount as measures. You have chosen to use 8 clusters and notice that 2 clusters only have 3 customers assigned. What should you do?

- A. Decrease the number of measures used
- B. Increase the number of clusters
- C. Decrease the number of clusters
- D. Identify additional measures to add to the analysis

#### Correct Answer: C

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**

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 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.

# **QUESTION 5**

You are working as a data science consultant for a gaming company. You have three member team and all other stake holders are from the company itself like project managers and project sponsored, data team etc. During the discussion project managed asked you that when can you tell me that the model you are using is robust enough, after which step you can consider answer for this question?

- A. Data Preparation
- B. Discovery
- C. Operationalize
- D. Model planning
- E. Model building
- Correct Answer: E

To answer whether the model you are building is robust enough or not you need to have answer below questions at least

-Model is performing as expected with the test data or not?

-Whatever hypothesis defined in the initial phase is being tested or not?

-Do we need more data?

- Domain experts are convinced or not with the model? And all these can be answered when you have built the model and tested with the test data sets. Hence, correct option will be Model Building.

PROFESSIONAL-DATA-SCIENTIST Practice Test

DATABRICKS-CERTIFIED- DATABRICKS-CERTIFIED- DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

**PROFESSIONAL-DATA-SCIENTIST Exam** Questions