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.passapply.com/databricks-certified-professional-data-scientist.html

100% Passing Guarantee
100% Money Back Assurance

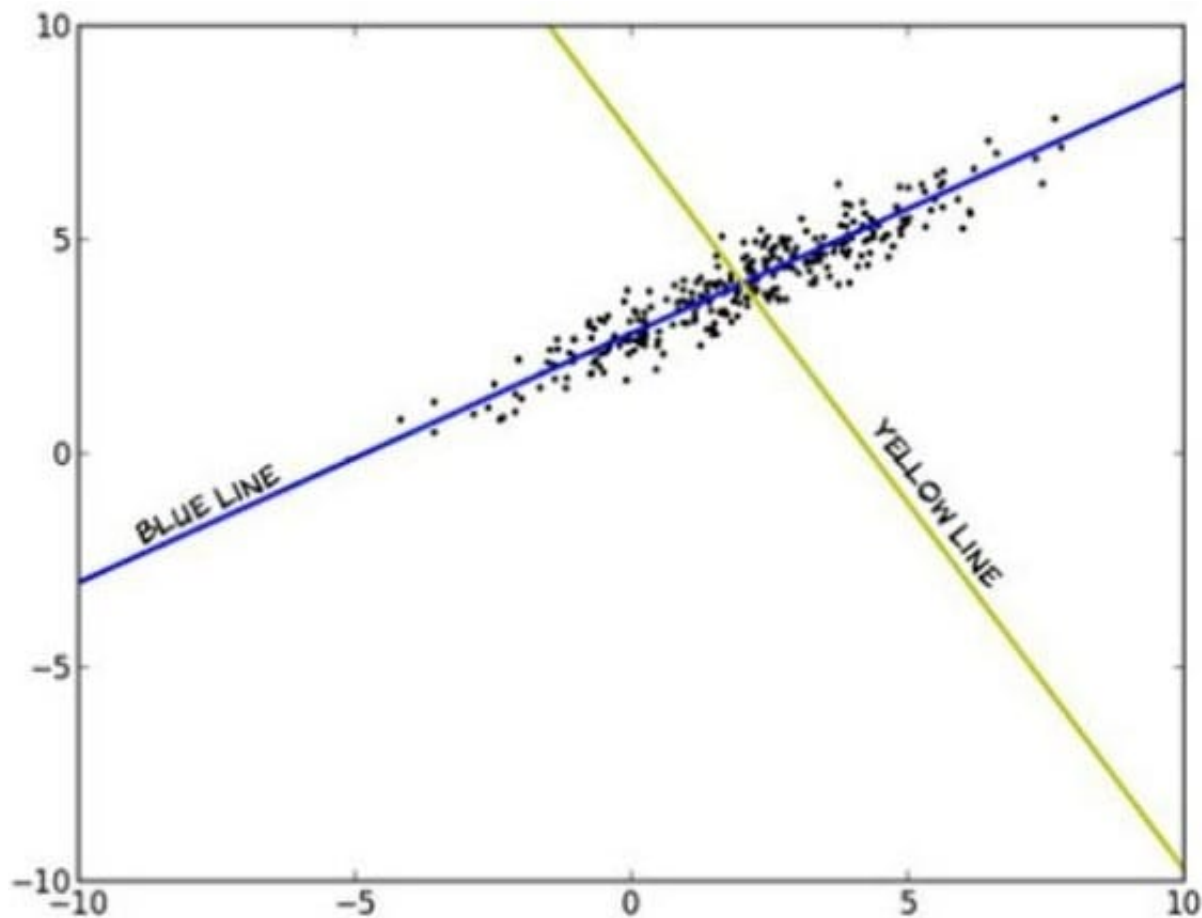Following Questions and Answers are all new published by Databricks Official Exam Center

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST Practice Test |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

2 / 5

QUESTION 1

The figure below shows a plot of the data of a data matrix M that is 1000 x 2. Which line represents the first principal component?

A. yellow

B. blue

C. Neither

Correct Answer: B

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The first principal component corresponds to the greatest variance in the data. The blue line is evidently this first principal component, because if we project the data onto the blue line, the data is more spread out (higher variance) than if projected onto any other line, including the yellow one.

QUESTION 2

Select the correct algorithm of unsupervised algorithm

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST Practice Test |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

3 / 5

A. K-Nearest Neighbors

B. K-Means

C. Support Vector Machines

D. Naive Bayes

Correct Answer: A

Explanation: Sup Supervised learning tasks Classification Regression k-Nearest Neighbors Linear Naive Bayes Locally weighted linear Support vector machines Ridge Decision trees Lasso Unsupervised learning tasks Clustering Density estimation k-Means Expectation maximization DBSCAN Parzen window

---

QUESTION 3

In unsupervised learning which statements correctly applies?

A. It does not have a target variable

B. Instead of telling the machine Predict Y for our data X, we\\'re asking What can you tell me about X?

C. telling the machine Predict Y for our data X

Correct Answer: AB

Explanation: In unsupervised learning we don\\'t have a target variable as we did in classification and regression.

Instead of telling the machine Predict Y for our data X, we\\'re asking What can you tell me about X?

Things we ask the machine to tell us about

X may be What are the six best groups we can make out of X? or What three features occur together most frequently in X?

QUESTION 4

Which of the following statement true with regards to Linear Regression Model?

A. Ordinary Least Square can be used to estimates the parameters in linear model

B. In Linear model, it tries to find multiple lines which can approximate the relationship between the outcome and input variables.

C. Ordinary Least Square is a sum of the individual distance between each point and the fitted line of regression model.

D. Ordinary Least Square is a sum of the squared individual distance between each point and the fitted line of regression model.

Correct Answer: AD

Explanation: Linear regression model are represented using the below equation

---

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST Practice Test |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

4 / 5

$$Y=B(0) + B(1)X$$

Where B(0) is intercept and B(1) is a slope. As B(0) and B(1) changes then fitted line also shifts accordingly on the plot. The purpose of the Ordinary Least Square method is to estimates these parameters B(0) and B(1). And similarly it is a sum of squared distance between the observed point and the fitted line. Ordinary least squares (OLS) regression minimizes the sum of the squared residuals. A model fits the data well if the differences between the observed values and the model\\'s predicted values are small and unbiased.

QUESTION 5

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

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST Practice Test |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

5 / 5