

# DATABRICKS-CERTIFIED- PR OFESIONAL-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.pass2lead.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

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



### QUESTION 1

What type of output generated in case of linear regression?

- A. Continuous variable
- B. Discrete Variable
- C. Any of the Continuous and Discrete variable
- D. Values between 0 and 1

Correct Answer: A

Explanation: Linear regression model generate continuous output variable.

---

### QUESTION 2

Regularization is a very important technique in machine learning to prevent over fitting. And Optimizing with a L1 regularization term is harder than with an L2 regularization term because

- A. The penalty term is not differentiate
- B. The second derivative is not constant
- C. The objective function is not convex
- D. The constraints are quadratic

Correct Answer: A

Explanation: Regularization is a very important technique in machine learning to prevent overfitting. Mathematically speaking, it adds a regularization term in order to prevent the coefficients to fit so perfectly to overfit. The difference between the L1 and L2 is just that L2 is the sum of the square of the weights, while L1 is just the sum of the weights. Much of optimization theory has historically focused on convex loss functions because they're much easier to optimize than non-convex functions: a convex function over a bounded domain is guaranteed to have a minimum, and it's easy to find that minimum by following the gradient of the function at each point no matter where you start. For non-convex functions, on the other hand, where you start matters a great deal; if you start in a bad position and follow the gradient, you're likely to end up in a local minimum that is not necessarily equal to the global minimum. You can think of convex functions as cereal bowls: anywhere you start in the cereal bowl, you're likely to roll down to the bottom. A non-convex function is more like a skate park: lots of ramps, dips, ups and downs. It's a lot harder to find the lowest point in a skate park than it is a cereal bowl.

---

### QUESTION 3

What are the advantages of the mutual information over the Pearson correlation for text classification problems?

- A. The mutual information has a meaningful test for statistical significance.
- B. The mutual information can signal non-linear relationships between the dependent and independent variables.

- C. The mutual information is easier to parallelize.
- D. The mutual information doesn't assume that the variables are normally distributed.

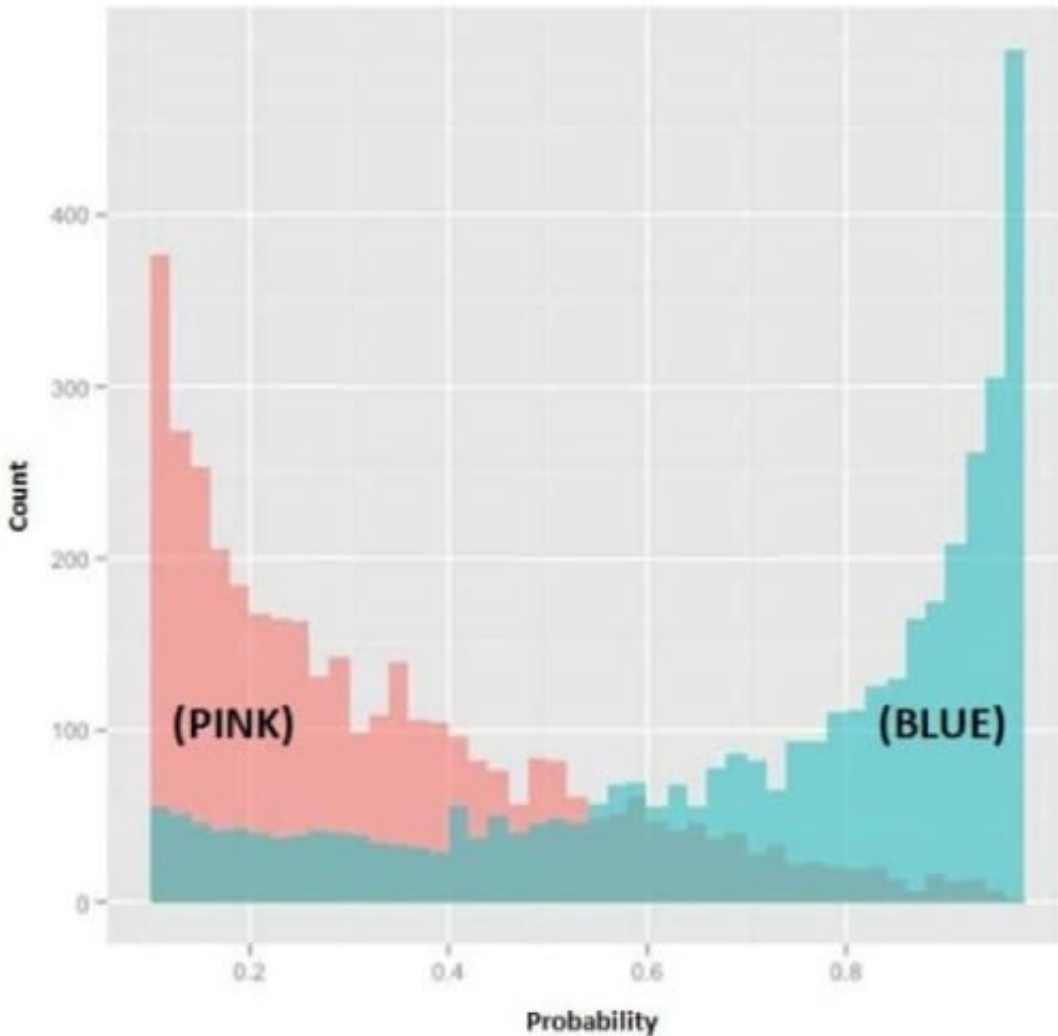
Correct Answer: C

Explanation: A linear scaling of the input variables (that may be caused by a change of units for the measurements) is sufficient to modify the PCA results. Feature selection methods that are sufficient for simple distributions of the patterns belonging to different classes can fail in classification tasks with complex decision boundaries. In addition, methods based on a linear dependence (like the correlation) cannot take care of arbitrary relations between the pattern coordinates and the different classes. On the contrary, the mutual information can measure arbitrary relations between variables and it does not depend on transformations acting on the different variables. This item concerns itself with feature selection for a text classification problem and references mutual information criteria. Mutual information is a bit more sophisticated than just selecting based on the simple correlation of two numbers because it can detect non-linear relationships that will not be identified by the correlation. Whenever possible: mutual information is a better feature selection technique than correlation. Mutual information is a quantification of the dependency between random variables. It is sometimes contrasted with linear correlation since mutual information captures nonlinear dependence. Correlation analysis provides a quantitative means of measuring the strength of a linear relationship between two vectors of data. Mutual information is essentially the measure of how much "knowledge" one can gain of a certain variable by knowing the value of another variable.

---

#### QUESTION 4

Refer to Exhibit



In the exhibit, the x-axis represents the derived probability of a borrower defaulting on a loan. Also in the exhibit, the pink represents borrowers that are known to have not defaulted on their loan, and the blue represents borrowers that are known to have defaulted on their loan. Which analytical method could produce the probabilities needed to build this exhibit?

- A. Linear Regression
- B. Logistic Regression
- C. Discriminant Analysis
- D. Association Rules

Correct Answer: B

### QUESTION 5

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

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

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide](#)

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Exam Questions](#)