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

QUESTION 1

Which of the following question statement falls under data science category?

A. What happened in last six months?

B. How many products have been sold in a last month?

C. Where is a problem for sales?

D. Which is the optimal scenario for selling this product?

E. What happens, if these scenario continues?

Correct Answer: DE

Explanation: This question wants to check your understanding about BI and Data Science. BI was already existing and analytics team already using it. They need to improve and learn data science technique to solve some problems. If you check the option given in the question, it will confuse you. But if you have worked in BI or as a Data Scientist then it is easy to answer. First 3 option can be easily answered using reporting solution, what sales happened in last six month, what was the problem etc. But for the last two option you need to apply data science techniques like which all scenarios are optimal for product sales, you need to collect the data and applying various techniques for that. Hence, last two option can only be answered using Data Science technique And for this you need to apply techniques like Optimization, predictive modeling, statistical analysis on structured and un-structured data.

QUESTION 2

A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. The response variable, admit/don\\'t admit, is a binary variable.

Above is an example of:

A. Linear Regression

B. Logistic Regression

C. Recommendation system

D. Maximum likelihood estimation

E. Hierarchical linear models

Correct Answer: B

Explanation: Logistic regression Pros: Computationally inexpensive, easy to implement, knowledge representation easy to interpret Cons: Prone to underfitting, may have low accuracy Works with: Numeric values, nominal values

QUESTION 3

Reducing the data from many features to a small number so that we can properly visualize it in two or three dimensions.

It is done in_____

A. supervised learning

B. un-supervised learning

C. k-Nearest Neighbors

D. Support vector machines

Correct Answer: B

Explanation: The opposite of supervised learning is a set of tasks known as unsupervised learning. In unsupervised learning, there\\'s no label or target value given for the data. A task where we group similar items together is known as clustering. In unsupervised learning, we may also want to find statistical values that describe the data. This is known as density estimation. Another task of unsupervised learning may be reducing the data from many features to a small number so that we can properly visualize it in two or three dimensions

QUESTION 4

You are working with the Clustering solution of the customer datasets. There are almost 40 variables are available for each customer and almost 1.00,0000 customer\\'s data is available. You want to reduce the number of variables for clustering, what would you do?

A. You will randomly reduce the number of variables

B. You will find the correlation among the variables and from their variables are not co- related will be discarded.

C. You will find the correlation among the variables and from the highly co-related variables, you will be considering only one or two variables from it.

D. You cannot discard any variable for creating clusters.

E. You can combine several variables in one variable

Correct Answer: CE

Explanation: When you are applying clustering technique and you find that there are quite a huge number of variables are available. Then it is better the find the co-relation among the variables and consider only one or two variables from the highly co-related variables. Because highly co-related variable will have the same effect, while creating the cluster. We can use scatter plot matrix among the variables to find the co-relation. You can also combine several variables into a single variable. For example if you have two values in the dataset like Asset and Debt than by combining these two values like Debt to Asset ratio and use it while creating the cluster.

QUESTION 5

Consider flipping a coin for which the probability of heads is p, where p is unknown, and our goa is to estimate p. The obvious approach is to count how many times the coin came up heads and divide by the total number of coin flips. If we flip the coin 1000 times and it comes up heads 367 times, it is very reasonable to estimate p as approximately 0.367. However, suppose we flip the coin only twice and we get heads both times. Is it reasonable to estimate p as 1.0? Intuitively, given that we only flipped the coin twice, it seems a bit rash to conclude that the coin will always come up heads, and_____is a way of avoiding such rash conclusions.

A. Naive Bayes

B. Laplace Smoothing

C. Logistic Regression

D. Linear Regression

Correct Answer: B

Explanation: Smooth the estimates: consider flipping a coin for which the probability of heads is p, where p is unknown, and our goal is to estimate p. The obvious approach is to count how many times the coin came up heads and divide by the total number of coin flips. If we flip the coin 1000 times and it comes up heads 367 times, it is very reasonable to estimate p as approximately 0.367. However, suppose we flip the coin only twice and we get heads both times. Is it reasonable to estimate p as 1.0? Intuitively, given that we only flipped the coin twice, it seems a bit rash to conclude that the coin will always come up heads, and smoothing is a way of avoiding such rash conclusions. A simple smoothing method, called Laplace smoothing (or Laplace\\'s law of succession or add-one smoothing in RandN), is to estimate p by (one plus the number of heads) / (two plus the total number of flips). Said differently, if we are keeping count of the number of heads and the number of tails, this rule is equivalent to starting each of our counts at one, rather than zero. Another advantage of Laplace smoothing is that it avoids estimating any probabilities to be zero, even for events never observed in the data. Laplace add-one smoothing now assigns too much probability to unseen words

QUESTION 6

You are working in a classification model for a book, written by HadoopExam Learning Resources and decided to use building a text classification model for determining whether this book is for Hadoop or Cloud computing. You have to select the proper features (feature selection) hence, to cut down on the size of the feature space, you will use the mutual information of each word with the label of hadoop or cloud to select the 1000 best features to use as input to a Naive Bayes model. When you compare the performance of a model built with the 250 best features to a model built with the 1000 best features, you notice that the model with only 250 features performs slightly better on our test data.

What would help you choose better features for your model?

A. Include least mutual information with other selected features as a feature selection criterion

B. Include the number of times each of the words appears in the book in your model

C. Decrease the size of our training data

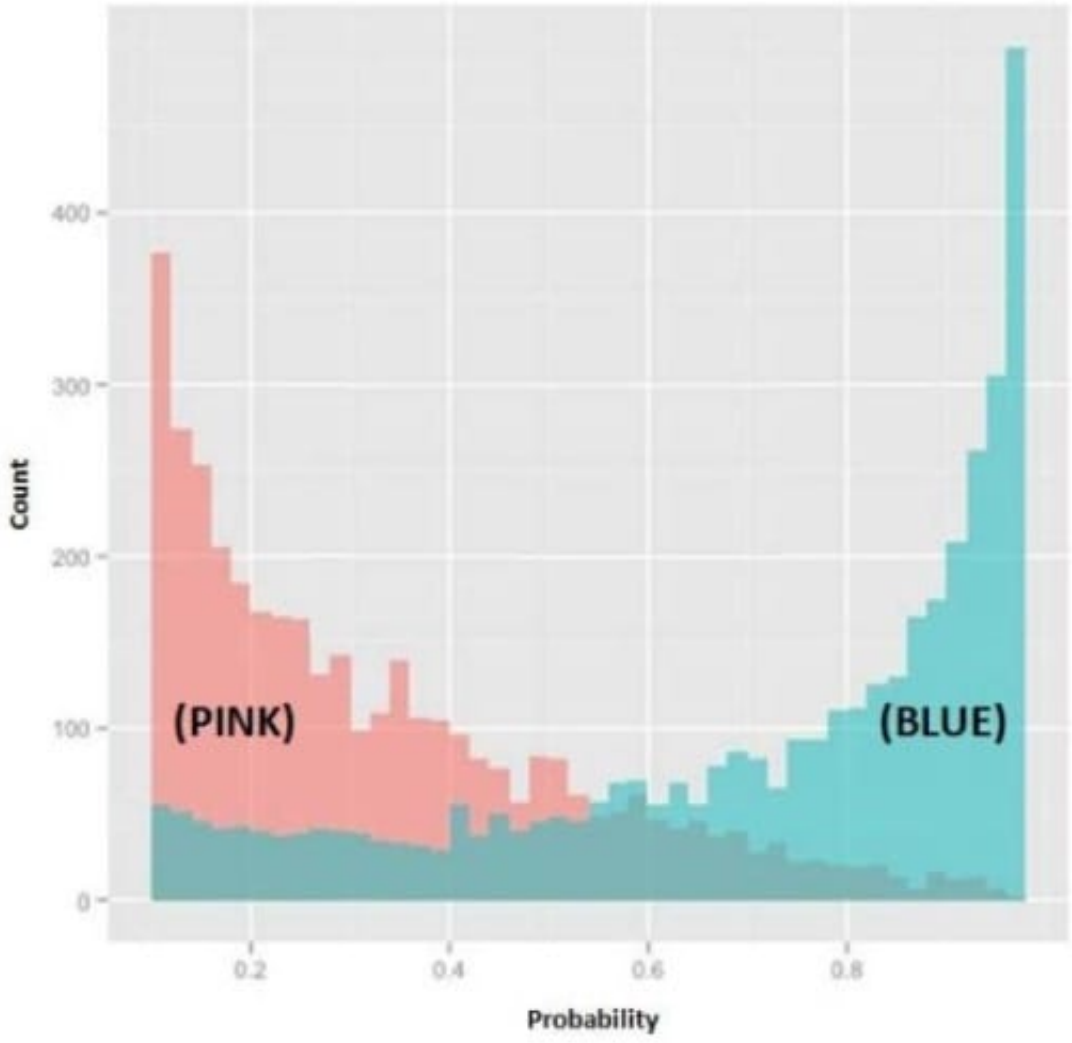
D. Evaluate a model that only includes the top 100 words

Correct Answer: A

Explanation: Correlation measures the linear relationship (Pearson\\'s correlation) or monotonic relationship (Spearman\\'s correlation) between two variables, X and Y. Mutual information is more general and measures the reduction of uncertainty in Y after observing X. It is the KL distance between the joint density and the product of the individual densities. So MI can measure non-monotonic relationships and other more complicated relationships Mutual information is a quantification of the dependency between random variables. It is sometimes contrasted with linear correlation since mutual information captures nonlinear dependence. Features with high mutual information with the predicted value are good. However a feature may have high mutual information because it is highly correlated with another feature that has already been selected. Choosing another feature with somewhat less mutual information with the predicted value, but low mutual information with other selected features, may be more beneficial. Hence it may help to also prefer features that are less redundant with other selected features.

QUESTION 7

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 8

A problem statement is given as below

Hospital records show that of patients suffering from a certain disease, 75% die of it. What is the probability that of 6 randomly selected patients, 4 will recover?

Which of the following model will you use to solve it.

A. Binomial

B. Poisson

C. Normal

D. Any of the above

Correct Answer: A

QUESTION 9

In which of the following scenario we can use naTve Bayes theorem for classification

A. Classify whether a given person is a male or a female based on the measured features. The features include height, weight and foot size.

B. To classify whether an email is spam or not spam

C. To identify whether a fruit is an orange or not based on features like diameter, color and shape

Correct Answer: ABC

Explanation: naive Bayes classifiers have worked quite well in many real-world situations, famously document classification and spam filtering. They requires a small amount of training data to estimate the necessary parameters

QUESTION 10

You are working in an ecommerce organization, where you are designing and evaluating a recommender system, you need to select which of the following metric wilt always have the largest value?

A. Root Mean Square Error

B. Sum of Errors

C. Mean Absolute Error

D. Both land 2

E. Information is not good enough.

Correct Answer: E

QUESTION 11

Suppose you have been given two Random Variables X and Y, whose joint distribution is already known, the marginal distribution of X is simply the probability distribution of X averaging over information about Y. It is the probability distribution of X when the value of Y is not known. So how do you calculate the marginal distribution of X

A. This is typically calculated by summing the joint probability distribution over Y.

B. This is typically calculated by integrating the joint probability distribution over Y

C. This is typically calculated by summing (In case of discrete variable) the joint probability distribution over Y

D. This is typically calculated by integrating(In case of continuous variable) the joint probability distribution over Y.

Correct Answer: ABCD

Explanation: Given two random variables X and Y whose joint distribution is known, the marginal distribution of X is simply the probability distribution of X averaging over information about Y. It is the probability distribution of X when the value of Y is not known. This is typically calculated by summing or integrating the joint probability distribution over Y. \\' For discrete random variables, the marginal probability mass function can be written as Pr(X = x). This is

$$\Pr(X = x) = \sum_y \Pr(X = x, Y = y) = \sum_y \Pr(X = x | Y = y) \Pr(Y = y),$$

Text

Description automatically generated with low confidence where Pr(X = x,Y = y) is the joint distribution of X and Y, while Pr(X = x|Y = y) is the conditional distribution of X given Y In this case, the variable Y has been marginalized out. Bivariate marginal and joint probabilities for discrete random variables are often displayed as two-way tables. Similarly for continuous random variables, the marginal probability density function can be written as pX(x). This is

$$p_X(x) = \int_y p_{X,Y}(x, y)\ \mathrm{d}y = \int_y p_{X|Y}(x|y)\, p_Y(y)\ \mathrm{d}y,$$

Diagram Description automatically generated with medium confidence where pX.Y(x.y) gives the joint distribution of X and Y while pX|Y(x|y) gives the conditional distribution for X given Y Again: the variable Y has been marginalized out.

Note that a marginal probability can always be written as an expected value:

$$p_X(x) = \int_y p_{X|Y}(x|y)\, p_Y(y)\ \mathrm{d}y = \mathbb{E}_Y[p_{X|Y}(x|y)]$$

Text, letter Intuitively, the marginal probability of X is computed by examining the conditional probability of X given a particular value of Y, and then averaging this conditional probability over the distribution of all values of Y This follows from the definition of expected value, i.e. in general

$$\mathbb{E}_Y[f(Y)] = \int_y f(y) p_Y(y)\ \mathrm{d}y$$

A picture containing diagram

QUESTION 12

You are creating a Classification process where input is the income, education and current debt of a customer, what could be the possible output of this process?

A. Probability of the customer default on loan repayment

B. Percentage of the customer loan repayment capability

C. Percentage of the customer should be given loan or not

D. The output might be a risk class, such as "good", "acceptable", "average", or "unacceptable".

Correct Answer: D

Explanation: Classification is the process of using several inputs to produce one or more outputs. For example the input might be the income, education and current debt of a customer The output might be a risk class, such as "good", "acceptable", "average", or "unacceptable". Contrast this to regression where the output is a number not a class.

QUESTION 13

Which of the below best describe the Principal component analysis

A. Dimensionality reduction

B. Collaborative filtering

C. Classification

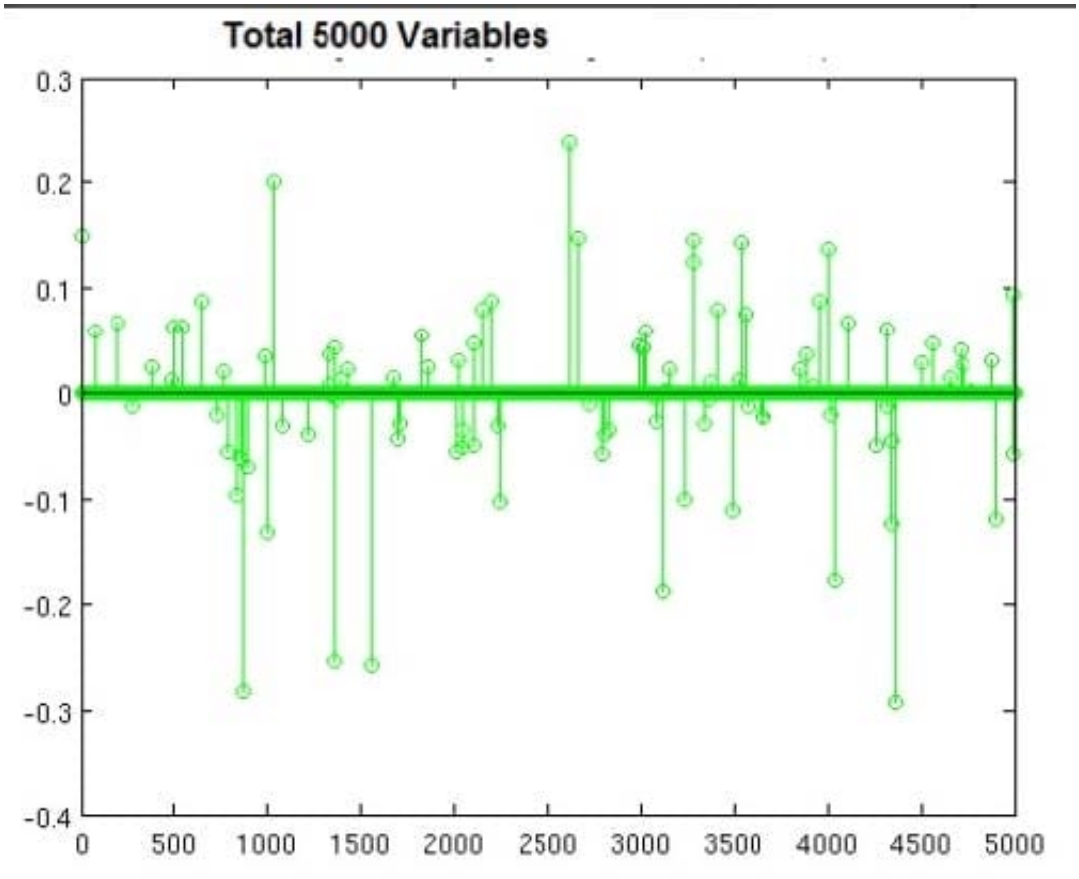
D. Regression

E. Clustering

Correct Answer: A

QUESTION 14

You are building a classifier off of a very high-dimensiona data set similar to shown in the image with 5000 variables (lots of columns, not that many rows). It can handle both dense and sparse input. Which technique is most suitable, and why?



A. Logistic regression with L1 regularization, to prevent overfitting

B. Naive Bayes, because Bayesian methods act as regularlizers

C. k-nearest neighbors, because it uses local neighborhoods to classify examples

D. Random forest because it is an ensemble method

Correct Answer: A

Explanation: Logistic regression is widely used in machine learning for classification problems. It is well-known that regularization is required to avoid over-fitting, especially when there is a only small number of training examples, or when there are a large number of parameters to be learned. In particular L1 regularized logistic regression is often used for feature selection, and has been shown to have good generalization performance in the presence of many irrelevant features. (Ng 2004; Goodman 2004) Unregularized logistic regression is an unconstrained convex optimization problem with a continuously differentiate objective function. As a consequence, it can be solved fairly efficiently with standard convex optimization methods, such as Newton\\'s method or conjugate gradient. However, adding the L1 regularization makes the optimization problem com-putationally more expensive to solve. If the L1 regulariza-tion is enforced by an L1 norm constraint on the parameLogistic regression is a classifier and L1 regularization tends to produce models that ignore dimensions of the input that are not predictive. This is particularly useful when the input contains many dimensions, k-nearest neighbors classification is also a classification technique, but relies on notions of distance. In a high- dimensional space, most every data point is "far" from others (the curse of dimensionality) and so these techniques break down. Naive Bayes is not inherently regularizing. Random forests represent an ensemble method; but an ensemble method is not necessarily more suitable to high-dimensional data. Practically, I think the biggest reasons for regularization are 1) to avoid overfitting by not generating high coefficients for predictors that are sparse. 2) to stabilize the estimates especially when there\\'s collinearity in the data.

1) is inherent in the regularization framework. Since there are two forces pulling each other in the objective function, if

there\\'s no meaningful loss reduction, the increased penalty from the regularization term wouldn\\'t improve the overall objective function. This is a great property since a lot of noise would be automatically filtered out from the model. To give you an example for 2), if you have two predictors that have same values, if you just run a regression algorithm on it since the data matrix is singular your beta coefficients will be Inf if you try to do a straight matrix inversion. But if you add a very small regularization lambda to it, you will get stable beta coefficients with the coefficient values evenly divided between the equivalent two variables. For the difference between L1 and L2, the following graph demonstrates why people bother to have L1 since L2 has such an elegant analytical solution and is so computationally straightforward. Regularized regression can also be represented as a constrained regression problem (since they are Lagrangian equivalent). The implication of this is that the L1 regularization gives you sparse estimates. Namely, in a high dimensional space, you got mostly zeros and a small number of non-zero coefficients. This is huge since it incorporates variable selection to the modeling problem. In addition, if you have to score a large sample with your model, you can have a lot of computational savings since you don\\'t have to compute features(predictors) whose coefficient is 0. I personally think L1 regularization is one of the most beautiful things in machine learning and convex optimization. It is indeed widely used in bioinformatics and large scale machine learning for companies like Facebook, Yahoo, Google and Microsoft.

QUESTION 15

You are working on a email spam filtering assignment, while working on this you find there is new word e.g. HadoopExam comes in email, and in your solutions you never come across this word before, hence probability of this words is coming in either email could be zero. So which of the following algorithm can help you to avoid zero probability?

A. Naive Bayes

B. Laplace Smoothing

C. Logistic Regression

D. All of the above

Correct Answer: B

Explanation: Laplace smoothing is a technique for parameter estimation which accounts for unobserved events. It is more robust and will not fail completely when data that has never been observed in training shows up.