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.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

QUESTION 1

Classification and regression are examples of_____.

A. supervised learning

B. un-supervised learning

C. Clustering

D. Density estimation

Correct Answer: A

Explanation: In classification, our job is to predict what class an instance of data should fall into. Another task in machine learning is regression. Regression is the prediction of a numeric value. Most people have probably seen an example of regression with a best-fit line drawn through some data points to generalize the data points. Classification and regression are examples of supervised learning. This set of problems is known as supervised because we\\'re telling the algorithm what to predict.

QUESTION 2

You are asked to create a model to predict the total number of monthly subscribers for a specific magazine. You are provided with 1 year\\'s worth of subscription and payment data, user demographic data, and 10 years worth of content of the magazine (articles and pictures). Which algorithm is the most appropriate for building a predictive model for subscribers?

A. Linear regression

B. Logistic regression

C. Decision trees

D. TF-IDF

Correct Answer: A

: A data model explicitly describes a relationship between predictor and response variables. Linear regression fits a data model that is linear in the model coefficients. The most common type of linear regression is a least-squares fit, which can fit both lines and polynomials, among other linear models. Before you model the relationship between pairs of quantities, it is a good idea to perform correlation analysis to establish if a linear relationship exists between these quantities. Be aware that variables can have nonlinear relationships, which correlation analysis cannot detect. For more information, see Linear Correlation. If you need to fit data with a nonlinear model, transform the variables to make the relationship linear. Alternatively try to fit a nonlinear function directly using either the Statistics and Machine Learning Toolbox nlinfit function, the Optimization Toolbox lsqcurvefit function, or by applying functions in the Curve Fitting Toolbox.

QUESTION 3

In which of the following scenario you should apply the Bay\\'s Theorem?

A. The sample space is partitioned into a set of mutually exclusive events {A1, A2, . .., An }.

B. Within the sample space, there exists an event B, for which P(B) > 0.

C. The analytical goal is to compute a conditional probability of the form: P(Ak | B).

D. In all above cases

Correct Answer: D

QUESTION 4

Refer to the exhibit.

Attribute	Info-Gain
Age	0.0310
Income	0.0100
Gender	0.0034
Credit Score	0.0456

You are building a decision tree. In this exhibit, four variables are listed with their respective values of info-gain. Based on this information, on which attribute would you expect the next split to be in the decision tree?

A. Credit Score

B. Age

C. Income

D. Gender

Correct Answer: A

QUESTION 5

Your company has organized an online campaign for feedback on product quality and you have all the responses for the product reviews, in the response form people have check box as well as text field. Now you know that people who do not fill in or write non-dictionary word in the text field are not considered valid feedback. People who fill in text field with proper English words are considered valid response. Which of the following method you should not use to identify whether the response is valid or not?

A. Naive Bayes

B. Logistic Regression

C. Random Decision Forests

D. Any one of the above

Correct Answer: D

In this problem you have been given high-dimensional independent variables like yeS; nO; no English words , 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