

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

- Instant Download After Purchase
- 100% Money Back Guarantee
- 😳 365 Days Free Update

VCE & PDF

PassApply.com

800,000+ Satisfied Customers





#### **QUESTION 1**

Spam filtering of the emails is an example of

- A. Supervised learning
- B. Unsupervised learning
- C. Clustering
- D. 1 and 3 are correct
- E. 2 and 3 are correct

Correct Answer: A

Explanation: Clustering is an example of unsupervised learning. The clustering algorithm finds groups within the data without being told what to look for upfront. This contrasts with classification, an example of supervised machine learning, which is the process of determining to which class an observation belongs. A common application of classification is spam filtering. With spam filtering we use labeled data to train the classifier: e-mails marked as spam or ham.

#### **QUESTION 2**

Select the correct algorithm of unsupervised algorithm

- 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**

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 4**

Select the correct objectives of principal component analysis:

#### A. To reduce the dimensionality of the data set



- B. To identify new meaningful underlying variables
- C. To discover the dimensionality of the data set

D. Only 1 and 2

E. All 1, 2 and 3

Correct Answer: E

Explanation: 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. Objectives of principal component analysis

1.

To discover or to reduce the dimensionality of the data set.

2.

To identify new meaningful underlying variables.

#### **QUESTION 5**

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

### PROFESSIONAL-DATA-SCIENTIST VCE Dumps

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

**PROFESSIONAL-DATA-SCIENTIST Exam** Questions