

DATABRICKS-CERTIFIED-PR OFESSIONAL-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 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

You are working on a Data Science project and during the project you have been gibe a responsibility to interview all the stakeholders in the project. In which phase of the project you are?

- A. Discovery
- **B.** Data Preparations
- C. Creating Models
- **D. Executing Models**
- E. Creating visuals from the outcome
- F. Operationnalise the models

Correct Answer: A

Explanation: During the discovery phase you will be interviewing all the project stakeholders because they would be having quite a good amount of knowledge for the problem domain you will be working and you also interviewing project sponsors you will get to know what all are the expectations once project get completed. Hence, you will be noting down all the expectations from the project as well as you will be using their expertise in the domain.

QUESTION 2

Logistic regression is a model used for prediction of the probability of occurrence of an event. It makes use of several variables that may be.....

A. Numerical

- B. Categorical
- C. Both 1 and 2 are correct
- D. None of the 1 and 2 are correct

Correct Answer: C

Explanation: Logistic regression is a model used for prediction of the probability of occurrence of an event. It makes use of several predictor variables that may be either numerical or categories.

QUESTION 3

Suppose that the probability that a pedestrian will be tul by a car while crossing the toad at a pedestrian crossing without paying attention to the traffic light is lo be computed. Let H be a discrete random variable taking one value from (Hit. Not Hit). Let L be a discrete random variable taking one value from (Red. Yellow. Green).

Realistically, H will be dependent on L That is, P(H = Hit) and P(H = Not Hit) will take different values depending on whether L is red, yellow or green. A person is. for example, far more likely to be hit by a car when trying to cross while Hie lights for cross traffic are green than if they are red In other words, for any given possible pair of values for Hand L.



one must consider the joint probability distribution of H and L to find the probability* of that pair of events occurring together if Hie pedestrian ignores the state of the light

Here is a table showing the conditional probabilities of being bit. defending on ibe stale of the lights (Note that the columns in this table must add up to 1 because the probability of being hit oi not hit is 1 regardless of the stale of the light.)

Conditional distribution: P(H L)				
	L=Green	L=Yellow	L=Red	
H=Not Hit	0.99	0.9	0.2	
H=Hit	0.01	0.1	0.8	

To find the joint probability distribution, we need more data. Let's say that P(L=green) = 0.2, P(L=yellow) = 0.1, and P(L=red) = 0.7. Multiplying each column in the conditional distribution by the probability of that column occurring, we find the joint probability distribution of H and L, given in the central 2×3 block of entries. (Note that the cells in this 2×3 block add up to 1).

Join				
	L-Green	L-Yellow	L-Red	Marginal probability P(H)
H=Not Hit	0.198	0.09	0.14	0.428
H=Hit	0.002	0.01	0.56	0.572
Total	0.2	0.1	0.7	1

Select the correct statement which applies to above example

A. The marginal probability P(H=Hit) is the sum along the H=Hit row of this joint distribution table, as this is the probability of being hit when the lights are red OR yellow OR green.

B. marginal probability that P(H=Not Hit) is the sum of the H=Not Hit row

C. marginal probability that P(H=Not Hit) is the sum of the H= Hit row

Correct Answer: AB

Explanation: The marginal probability P(H=Hit) is the sum along the H=Hit row of this joint distribution table, as this is the probability of being hit when the lights are red OR yellow OR green. Similarly, the marginal probability that P(H=Not Hit) is the sum of the H=Not Hit row

QUESTION 4

You are using k-means clustering to classify heart patients for a hospital. You have chosen Patient Sex, Height, Weight, Age and Income as measures and have used 3 clusters. When you create a pair-wise plot of the clusters, you notice that there is significant overlap between the clusters. What should you do?

- A. Identify additional measures to add to the analysis
- B. Remove one of the measures
- C. Decrease the number of clusters
- D. Increase the number of clusters



Correct Answer: C

QUESTION 5

Projecting a multi-dimensional dataset onto which vector has the greatest variance?

- A. first principal component
- B. first eigenvector
- C. not enough information given to answer
- D. second eigenvector
- E. second principal component

Correct Answer: A

Explanation: The method based on principal component analysis (PCA) evaluates the features according to the projection of the largest eigenvector of the correlation matrix on the initial dimensions, the method based on Fisher\\'s linear discriminant analysis evaluates. Them according to the magnitude of the components of the discriminant vector. The first principal component corresponds to the greatest variance in the data, by definition. If we project the data onto the first principal component line, the data is more spread out (higher variance) than if projected onto any other line, including other principal components.

Latest DATABRICKS-CERT	DATABRICKS-CERTIFIED-	DATABRICKS-CERTIFIED-
IFIED-PROFESSIONAL-	PROFESSIONAL-DATA-	PROFESSIONAL-DATA-
DATA-SCIENTIST Dumps	SCIENTIST PDF Dumps	SCIENTIST Study Guide