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.pass4itsure.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

If you are trying to predict or forecast a discrete target value, then which is the correct options?

A. Supervised Learning regression algorithms

B. Supervised Learning classification algorithms

C. Un supervised Learning

D. Density estimation algorithm

Correct Answer: B

Explanation: If you\\'re trying to predict or forecast a target value, then you need to look into supervised learning. If not, then unsupervised learning is the place you want to be. If you\\'ve chosen supervised learning, what\\'s your target value? Is it

a discrete value like Yes/No:

1/2/3, A/B/C: or Red/Yellow/Black? If so: then you want to look into classification. If the target value can take on a number of values, say any value from 0.00 to 100.00: or-999 to 999, or +_to -_, then you need to look into regression.

QUESTION 2

Suppose a man told you he had a nice conversation with someone on the train. Not knowing anything about this conversation, the probability that he was speaking to a woman is 50% (assuming the train had an equal number of men and women and the speaker was as likely to strike up a conversation with a man as with a woman). Now suppose he also told you that his conversational partner had long hair. It is now more likely he was speaking to a woman, since women are more likely to have long hair than men._____ can be used to calculate the probability that the person was a woman.

A. SVM

B. MLE

C. Bayes\\' theorem

D. Logistic Regression

Correct Answer: C

Explanation: To see how this is done, let W represent the event that the conversation was held with a woman, and L denote the event that the conversation was held with a long-haired person. It can be assumed that women constitute half the

population for this example. So, not knowing anything else, the probability that W occurs is P(W) = 0.5. Suppose it is also known that 75% of women have long hair which we denote as P(L |W) = 0.75 (read: the probability of event L given

event W is 0.75, meaning that the probability of a person having long hair (event "L"): given that we already know that the person is a woman ("event W") is 75%). Likewise, suppose it is known that 15% of men have long hair, or P(L |M) =

0.15; where M is the complementary event of W: i.e.; the event that the conversation was held with a man (assuming that every human is either a man or a woman).

Our goal is to calculate the probability that the conversation was held with a woman, given the fact that the person had long hair, or, in our notation, P(W |L). Using the formula for Bayes\\' theorem, we have:

$$P(W|L) = \frac{P(L|W)P(W)}{P(L)} = \frac{P(L|W)P(W)}{P(L|W)P(W) + P(L|M)P(M)}$$

Text Description automatically generated with low confidence where we have used the law of total probability to expand P(L),

The numeric answer can be obtained by substituting the above values into this formula (the algebraic multiplication is annotated using " *", the centered dot). This yields

$$P(W|L) = \frac{0.75 \cdot 0.50}{0.75 \cdot 0.50 + 0.15 \cdot 0.50} = \frac{5}{6} \approx 0.83,$$

A picture containing table i.e., the probability that the conversation was held with a woman, given that the person had long hair is about 83%. More examples are provided below.

QUESTION 3

You are creating a regression model with the input income, education and current debt of a customer, what could be the possible output from this model?

A. Customer fit as a good

B. Customer fit as acceptable or average category

C. expressed as a percent, that the customer will default on a loan

D. 1 and 3 are correct

E. 2 and 3 are correct

Correct Answer: C

Explanation: Regression 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 the probability, expressed as a percent that the customer will default on a loan. Contrast this to classification where the output is not a number, but a class.

QUESTION 4

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 5

You have used k-means clustering to classify behavior of 100, 000 customers for a retail store. You decide to use household income, age, gender and yearly purchase amount as measures. You have chosen to use 8 clusters and notice that 2 clusters only have 3 customers assigned. What should you do?

A. Decrease the number of measures used

B. Increase the number of clusters

C. Decrease the number of clusters

D. Identify additional measures to add to the analysis

Correct Answer: C

Explanation: kmeans uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. This algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible. You can control the details of the minimization using several optional input parameters to kmeans, including ones for the initial values of the cluster centroids, and for the maximum number of iterations. Clustering is primarily an exploratory technique to discover hidden structures of the data: possibly as a prelude to more focused analysis or decision processes. Some specific applications of k-means are image processing^ medical and customer segmentation. Clustering is often used as a lead-in to classification. Once the clusters are identified, labels can be applied to each cluster to classify each group based on its characteristics. Marketing and sales groups use k-means to better identify customers who have similar behaviors and spending patterns.