DUMPS ARENA

Databricks Certified Professional Data Scientist Exam

Databricks Databricks-Certified-Professional-Data-Scientist

Version Demo

Total Demo Questions: 10

Total Premium Questions: 138

Buy Premium PDF

QUESTION NO: 1

Which of the following true with regards to the K-Means clustering algorithm?

A. Labels are not pre-assigned to each objects in the cluster.

B. Labels are pre-assigned to each objects in the cluster.

C. It classify the data based on the labels.

D. It discovers the center of each cluster.

E. It find each objects fall in which particular cluster

ANSWER: A D E

Explanation:

: Clustering does not require any predefined labels on the object, rather it consider the attributes on the object. Hence, option-B is out. Clustering is different than classification technique.

Hence you can discard the option-C as well. It does not use the pre-defined labels, hence it is called unsupervised learning and option-Ais correct. Main purpose of the Clustering technique is to determine the center of each Cluster and then find the distance from that center. If object is near the center than it would fall in that particular cluster. Hence, finally you will have group or clusters created and get to know that objects fall in which particular cluster.

QUESTION NO: 2

What is the considerable difference between L1 and L2 regularization?

A. L1 regularization has more accuracy of the resulting model

B. Size of the model can be much smaller in L1 regularization than that produced by L2regularization

C. L2-regularization can be of vital importance when the application is deployed in resource-tight environments such as cell-phones.

D. All of the above are correct

ANSWER: B

Explanation:

: The two most common regularization methods are called L1 and L2 regularization. L1 regularization penalizes the weight vector for its L1-norm (i.e. the sum of the absolute values of the weights), whereas L2 regularization uses its L2-norm. There is usually not a considerable difference between the two methods in terms of the accuracy of the resulting model (Gao et al

2007), but L1 regularization has a significant advantage in practice. Because many of the weights of the features become zero as a result of L1regularized training, the size of the model can be much smaller than that produced by L2regularization. Compact models require less space on memory and storage, and enable the application to start up quickly. These merits can be of vital importance when the application is deployed in resource-tight environments such as cell-phones.

Regularization works by adding the penalty associated with the coefficient values to the error of the hypothesis. This way, an accurate hypothesis with unlikely coefficients would be penalized whila a somewhat less accurate but more conservative hypothesis with low coefficients would not be penalized as much.

81

QUESTION NO: 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).

Joint distribution: P(H,L)				
	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

ANSWER: A B

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 NO: 4

RMSE measures error of a predicted

A. Numerical Value

B. Categorical values

C. For booth Numerical and categorical values

ANSWER: A

QUESTION NO: 5

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.

E. '
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

Description automatically generated
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

Description automatically generated

ANSWER: A B C D

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

Description automatically generated

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

Description automatically generated

QUESTION NO: 6

What are the advantages of the Hashing Features?

A. Requires the less memory

B. Less pass through the training data

C. Easily reverse engineer vectors to determine which original feature mapped to a vector location

ANSWER: A B

Explanation:

: SGD-based classifiers avoid the need to predetermine vector size by simply picking a reasonable size and shoehorning the training data into vectors of that size. This approach is known as feature hashing. The shoehorning is done by picking one or more locations by using a hash of the name of the variable for continuous variables or a hash of the variable name and the category name or word for categorical, text-like, or word-like data.

This hashed feature approach has the distinct advantage of requiring less memory and one less pass through the training data, but it can make it much harder to reverse engineer vectors to determine which original feature mapped to a vector location. This is because multiple features may hash to the same location. With large vectors or with multiple locations per feature, this isn't a problem for accuracy but it can make it hard to understand what a classifier is doing.

An additional benefit of feature hashing is that the unknown and unbounded vocabularies typical of word-like variables aren't a problem.

QUESTION NO: 7

Select the correct statement which applies to K-Nearest Neighbors

A. No Assumption about the data

B. Computationally expensive

C. Require less memory

D. Works with Numeric Values

ANSWER: A B D

Explanation:

:

: k-Nearest Neighbors

Pros: High accuracy insensitive to outliers, no assumptions about data

Cons: Computationally expensive, requires a lot of memory

Works with: Numeric values, nominal values

QUESTION NO: 8

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

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 NO: 9

Which analytical method is considered unsupervised?

y1, y2, y3, ... , yn-1, yn

may have a trend component that is quadratic in nature. Which pattern of data will indicate that the trend in the time series data is quadratic in nature?

A. Naive Bayesian classifier

B. Decision tree

C. Linear regression

D. K-means clustering

ANSWER: D

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.

QUESTION NO: 10

Of all the smokers in a particular district, 40% prefer brand A and 60% prefer brand B. Of those smokers who prefer brand A. 30% are females, and of those who prefer brand B. 40% are female. What is the probability that a randomly selected smoker prefers brand A, given that the person selected is a female?

Which of the following is a best way to solve this problem?

A. Bays Theorem

B. Poisson Distribution

C. Binomial Distribution

D. None of the above

ANSWER: A