Databricks.Databricks-Certified-Professional-Data-Scientist.v2022-04-10.q51

Exam Code:	Databricks-Certified-Professional-Data-Scientist
Exam Name:	Databricks Certified Professional Data Scientist Exam
Certification Provider:	Databricks
Free Question Number:	51
Version:	v2022-04-10
# of views:	1870
# of Questions views:	1979
https://www.dumpsfiles.com/files/Databricks/Databricks-Certified-Professional-Data-Scientist/Databricks.Databricks-Certified-Professional-Data-Scientist.v2022-04-10.q51	

NEW QUESTION: 1
Select the sequence of the developing machine learning applications
A) Analyze the input data
B) Prepare the input data
C) Collect data
D) Train the algorithm
E) Test the algorithm
F) Use It
A. A, B, C, D, E, F
B. C, B, A, D, E, F
C. C, A, B, D, E, F
D. C, B, A, D, E, F
Answer: D (LEAVE A REPLY)
Explanation
1 Collect data. You could collect the samples by scraping a website and extracting data: or you could get information from an RSS feed or an API. You could have a device collect wind speed measurements and send them to you, or blood glucose levels, or anything you can measure. The number of options is endless. To save some time and effort you could use publicly available data
2 Prepare the input data. Once you have this data, you need to make sure it's in a useable format. The format we'll be using in this book is the Python list. We'll talk about Python more in a little bit, and lists are reviewed in appendix A.
The benefit of having this standard format is that you can mix and match algorithms and data sources. You may need to do some algorithm-specific formatting here. Some algorithms need features in a special format, some algorithms can deal with target variables and

features as strings, and some need them to be integers. We'll get to this later but the algorithm-specific formatting is usually trivial compared to collecting data.
3 Analyze the input data. This is looking at the data from the previous task. This could be as simple as looking at the data you've parsed in a text editor to make sure steps 1 and 2 are actually working and you don't have a bunch of empty values. You can also look at the data to see if you can recognize any patterns or if there's anything obvious^ such as a few data points that are vastly different from the rest of the set. Plotting data in one: two, or three dimensions can also help. But most of the time you'll have more than three features, and you can't easily plot the data across all features at one time. You could, however use some advanced methods we'll talk about later to distill multiple dimensions down to two or three so you can visualize the data.
4 If you're working with a production system and you know what the data should look like, or you trust its source: you can skip this step. This step takes human involvement, and for an automated system you don't want human involvement. The value of this step is that it makes you understand you don't have garbage coming in.
5 Train the algorithm. This is where the machine learning takes place. This step and the next step are where the "core" algorithms lie, depending on the algorithm.You feed the algorithm good clean data from the first two steps andextract knowledge or information. This knowledge you often store in a formatthat's readily useable by a machine for the next two steps.In the case of unsupervised learning, there's no training step because youdon't have a target value. Everything is used in the next step.
6 Test the algorithm. This is where the information learned in the previous step isput to use. When you're evaluating an algorithm, you'll test it to see how well itdoes. In the case of supervised learning, you have some known values you can use to evaluate the algorithm. In unsupervised learning, you may have to use some other metrics to evaluate the success. In either case, if you're not satisfied, you can go back to step 4, change some things, and try testing again. Often thecollection or preparation of the data may have been the problem, and you'll have to go back to step 1.
7 Use it. Here you make a real program to do some task, and once again you see if all the previous steps worked as you expected. You might encounter some new data and have to revisit steps 1-5.

NEW QUESTION: 2
Which of the following statement true with regards to Linear Regression Model?
A. Ordinary Least Square can be used to estimates the parameters in linear model
B. In Linear model, it tries to find multiple lines which can approximate the relationship between the outcome and input variables.
C. Ordinary Least Square is a sum of the individual distance between each point and the fitted line of regression model.
D. Ordinary Least Square is a sum of the squared individual distance between each point and the fitted line of regression model.

Answer: A,D (LEAVE A REPLY)

Explanation

Linear regression model are represented using the below equation

$$Y=B(0) + B(1)X$$

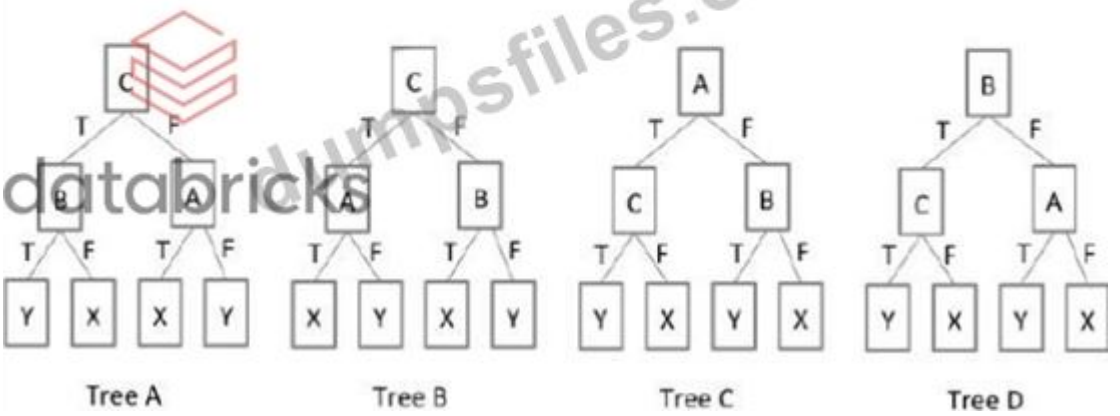
Where B(0) is intercept and B(1) is a slope. As B(0) and B(1) changes then fitted line also shifts accordingly on the plot. The purpose of the Ordinary Least Square method is to estimates these parameters B(0) and B(1).

And similarly it is a sum of squared distance between the observed point and the fitted line. Ordinary least squares (OLS) regression minimizes the sum of the squared residuals. A model fits the data well if the differences between the observed values and the model's predicted values are small and unbiased.

NEW QUESTION: 3

Refer to the Exhibit.

C	B	A	CLASS
T	T	T	X
T	T	F	Y
T	F	T	X
F	F	F	Y
F	T	T	X
F	F	T	Y



In the Exhibit, the table shows the values for the input Boolean attributes "A", "B", and "C". It also shows the values for the output attribute "class". Which decision tree is valid for the data?

A. Tree A

B. Tree D

C. Tree B

D. Tree C

Answer: C (LEAVE A REPLY)

NEW QUESTION: 4

Which is an example of supervised learning?

A. PCA

B. k-means clustering
C. SVD
D. EM
E. SVM
Answer: E (LEAVE A REPLY)
Explanation
SVMs can be used to solve various real world problems:
* SVMs are helpful in text and hypertext categorization as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.
* Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback.
* SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly.
* Hand-written characters can be recognized using SVM

NEW QUESTION: 5
You are working on a problem where you have to predict whether the claim is done valid or not. And you find that most of the claims which are having spelling errors as well as corrections in the manually filled claim forms compare to the honest claims. Which of the following technique is suitable to find out whether the claim is valid or not?
A. Naive Bayes
B. Logistic Regression
C. Random Decision Forests
D. Any one of the above
Answer: (SHOW ANSWER)
Explanation
In this problem you have been given high-dimensional independent variables like texts, corrections, 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

NEW QUESTION: 6
Question-26. There are 5000 different color balls, out of which 1200 are pink color. What is the maximum likelihood estimate for the proportion of "pink" items in the test set of color balls?
A. 2.4
B. 24 0
C. .24
D. .48

E. 4.8

Answer: C (LEAVE A REPLY)

Explanation

Given no additional information, the MLE for the probability of an item in the test set is exactly its frequency in the training set. The method of maximum likelihood corresponds to many well-known estimation methods in statistics. For example, one may be interested in the heights of adult female penguins, but be unable to measure the height of every single penguin in a population due to cost or time constraints. Assuming that the heights are normally (Gaussian) distributed with some unknown mean and variance, the mean and variance can be estimated with MLE while only knowing the heights of some sample of the overall population. MLE would accomplish this by taking the mean and variance as parameters and finding particular parametric values that make the observed results the most probable (given the model).

In general, for a fixed set of data and underlying statistical model the method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. Intuitively, this maximizes the "agreement" of the selected model with the observed data, and for discrete random variables it indeed maximizes the probability of the observed data under the resulting distribution. Maximum-likelihood estimation gives a unified approach to estimation, which is well-defined in the case of the normal distribution and many other problems. However in some complicated problems, difficulties do occur: in such problems, maximum-likelihood estimators are unsuitable or do not exist.

NEW QUESTION: 7

Which of the following could be features?

A. Words in the document

B. Symptoms of a diseases

C. Characteristics of an unidentified object

D. 0nly 1 and 2

E. All 1,2 and 3 are possible

Answer: (SHOW ANSWER)

Explanation

Any dataset that can be turned into lists of features. A feature is simply something that is either present or absent for a given item. In the case of documents, the features are the words in the document but they could also be characteristics of an unidentified object symptoms of a disease, or anything else that can be said to be present of absent.

NEW QUESTION: 8

A website is opened 3 times by a user. What is the probability of he clicks 2 times the advertisement, is best calculated by

A. Binomial

B. Poisson

C. Normal
D. Any of the above
Answer: A (LEAVE A REPLY)
Explanation
In a binomial distribution, only 2 parameters, namely n and p, are needed to determine the probability. Where p is the probability of success and q is the probability of failure in a binomial trial, then the expected number of successes in n trials.
This is a binomial distribution because there are only 2 possible outcomes (we get a 5 or we don't).

NEW QUESTION: 9
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
Answer: C (LEAVE A REPLY)
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:
Text Description automatically generated with low confidence

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

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 A picture containing table Description automatically generated

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

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.

NEW QUESTION: 10
Scenario: Suppose that Bob can decide to go to work by one of three modes of transportation, car, bus, or commuter train. Because of high traffic, if he decides to go by car. there is a 50% chance he will be late. If he goes by bus, which has special reserved lanes but is sometimes overcrowded, the probability of being late is only 20%. The commuter train is almost never late, with a probability of only 1 %, but is more expensive than the bus. Suppose that Bob is late one day, and his boss wishes to estimate the probability that he drove to work that day by car. Since he does not know Which mode of transportation Bob usually uses, he gives a prior probability of
1 3 to each of the three possibilities. Which of the following method the boss will use to estimate of the probability that Bob drove to work?
A. Naive Bayes
B. Linear regression
C. Random decision forests
D. None of the above
Answer: A (LEAVE A REPLY)
Explanation
Bayes' theorem (also known as Bayes' rule) is a useful tool for calculating conditional probabilities.

NEW QUESTION: 11
You are analyzing data in order to build a classifier model. You discover non-linear data and discontinuities that will affect the model. Which analytical method would you recommend?
A. Logistic Regression
B. Decision Trees

C. Linear Regression
D. ARIMA
Answer: (SHOW ANSWER)
Explanation
A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g.
whether a coin flip comes up heads or tails), each branch represents the outcome of the test and each leaf node represents a class label (decision taken after computing all attributes).
The paths from root to leaf represents classification rules.
In decision analysis a decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values (or expected utility) of competing alternatives are calculated.
A decision tree consists of 3 types of nodes:
1. Decision nodes - commonly represented by squares
2. Chance nodes - represented by circles
3. End nodes - represented by triangles
Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. If in practice decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities.
Decision trees, influence diagrams, utility functions, and other decision analysis tools and methods are taught to undergraduate students in schools of business, health economics, and public health, and are examples of operations research or management science methods.

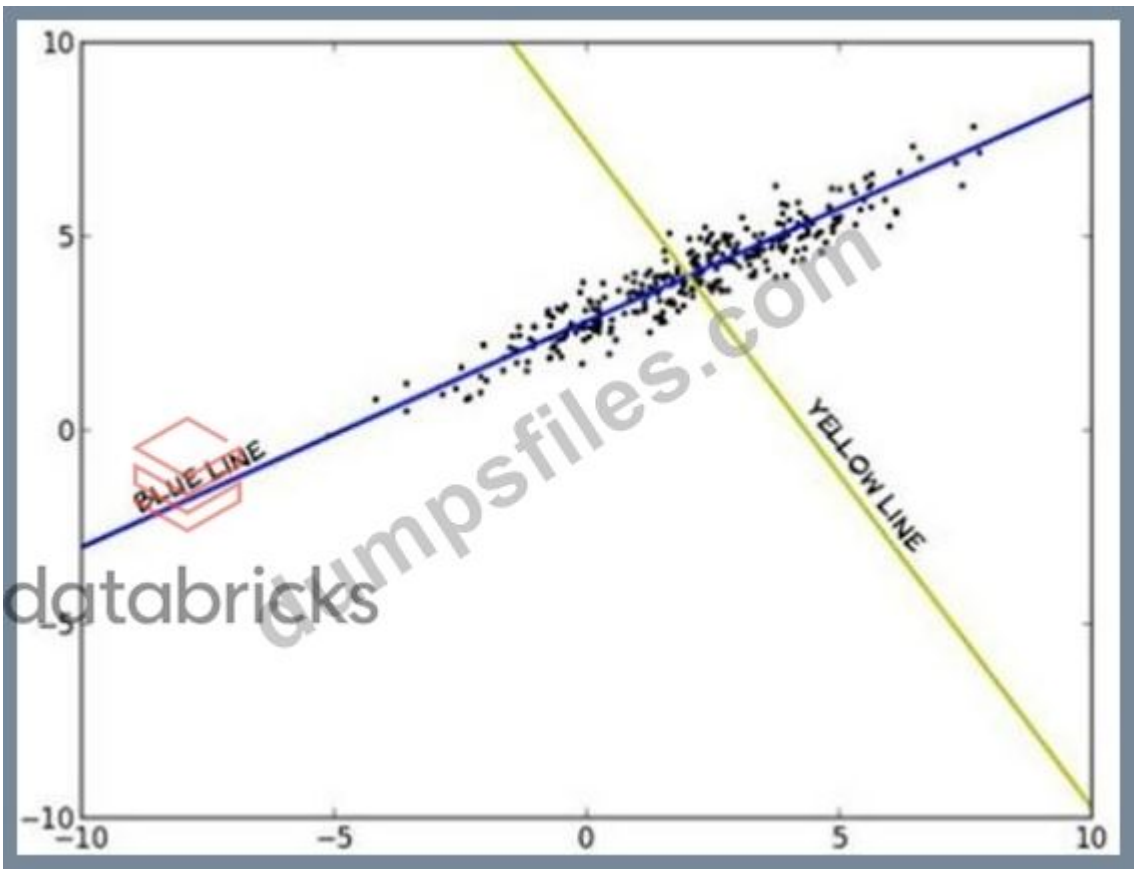
NEW QUESTION: 12
Question-3: In machine learning, feature hashing, also known as the hashing trick (by analogy to the kernel trick), is a fast and space-efficient way of vectorizing features (such as the words in a language), i.e., turning arbitrary features into indices in a vector or matrix. It works by applying a hash function to the features and using their hash values modulo the number of features as indices directly, rather than looking the indices up in an associative array. So what is the primary reason of the hashing trick for building classifiers?
A. It creates the smaller models
B. It requires the lesser memory to store the coefficients for the model
C. It reduces the non-significant features e.g. punctuations
D. Noisy features are removed
Answer: B (LEAVE A REPLY)
Explanation
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.
Models always have a coefficient per feature, which are stored in memory during model building. The hashing trick collapses a high number of features to a small number which reduces the number of coefficients and thus memory requirements. Noisy features are not removed; they are combined with other features and so still have an impact.
The validity of this approach depends a lot on the nature of the features and problem domain; knowledge of the domain is important to understand whether it is applicable or will likely produce poor results. While hashing features may produce a smaller model, it will be one built from odd combinations of real-world features, and so will be harder to interpret.
An additional benefit of feature hashing is that the unknown and unbounded vocabularies typical of word-like variables aren't a problem.

NEW QUESTION: 13
Regularization is a very important technique in machine learning to prevent overfitting. Mathematically speaking, it adds a regularization term in order to prevent the coefficients to fit so perfectly to overfit. The difference between the L1 and L2 is...
A. L2 is the sum of the square of the weights, while L1 is just the sum of the weights
B. L1 is the sum of the square of the weights, while L2 is just the sum of the weights
C. L1 gives Non-sparse output while L2 gives sparse outputs
D. None of the above
Answer: A (LEAVE A REPLY)
Explanation
Regularization is a very important technique in machine learning to prevent overfitting. Mathematically speaking, it adds a regularization term in order to prevent the coefficients to fit so perfectly to overfit. The difference between the L1 and L2 is just that L2 is the sum of the square of the weights, while L1 is just the sum of the weights. As follows: L1 regularization on least squares:
A picture containing text Description automatically generated

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_j \left(t(\mathbf{x}_j) - \sum_i w_i h_i(\mathbf{x}_j) \right)^2 + \lambda \sum_{i=1}^{k} w_i$$

NEW QUESTION: 14



The figure below shows a plot of the data of a data matrix M that is 1000 x 2. Which line represents the first principal component?

A. yellow

B. blue

C. Neither

Answer: B (LEAVE A REPLY)

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.

The first principal component corresponds to the greatest variance in the data. The blue line is evidently this first principal component, because if we project the data onto the blue line, the data is more spread out (higher variance) than if projected onto any other line, including the yellow one.

NEW QUESTION: 15

Which of the following are point estimation methods?

A. MAP

B. MLE

C. MMSE

Answer: A,B,C (LEAVE A REPLY)

Explanation
Point estimators
* minimum-variance mean-unbiased estimator (MVUE), minimizes the risk (expected loss) of the squared-error loss-function.
* best linear unbiased estimator (BLUE)
* minimum mean squared error (MMSE)
* median-unbiased estimator, minimizes the risk of the absolute-error loss function
* maximum likelihood (ML)
* method of moments, generalized method of moments

NEW QUESTION: 16
Digit recognition, is an example of.....
A. Classification
B. Clustering
C. Unsupervised learning
D. None of the above
Answer: A (LEAVE A REPLY)
Explanation
Supervised learning is fairly common in classification problems because the goal is often to get the computer to learn a classification system that we have created. Digit recognition: once again, is a common example of classification learning. More generally, classification learning is appropriate for any problem where deducing a classification is useful and the classification is easy to determine. In some cases, it might not even be necessary to give pre-determined classifications to every instance of a problem if the agent can work out the classifications for itself. This would be an example of unsupervised learning in a classification context.

NEW QUESTION: 17

In which phase of the data analytics lifecycle do Data Scientists spend the most time in a project?

A. Communicate Results

B. Discovery

C. Data Preparation

D. Model Building

Answer: C (LEAVE A REPLY)

NEW QUESTION: 18

Select the choice where Regression algorithms are not best fit

A. When the dimension of the object given

B. Weight of the person is given

C. Temperature in the atmosphere

D. Employee status

Answer: D (LEAVE A REPLY)

Explanation

Regression algorithms are usually employed when the data points are inherently numerical variables (such as the dimensions of an object the weight of a person, or the temperature in the atmosphere) but unlike Bayesian algorithms, they're not very good for categorical data (such as employee status or credit score description).

NEW QUESTION: 19

You are creating a model for the recommending the book at Amazon.com, so which of the following recommender system you will use you don't have cold start problem?

A. Naive Bayes classifier

B. Item-based collaborative filtering

C. User-based collaborative filtering

D. Content-based filtering

Answer: D (LEAVE A REPLY)

Explanation

The cold start problem is most prevalent in recommender systems. Recommender systems form a specific type of information filtering (IF) technique that attempts to present information items (movies, music, books, news, images, web pages) that are likely of interest to the user. Typically, a recommender system compares the user's profile to some reference characteristics. These characteristics may be from the information item (the content-based approach) or the user's social environment (the collaborative filtering approach). In the content-based approach, the system must be capable of matching the characteristics of an item against relevant features in the user's profile. In order to do this, it must first construct a sufficiently-detailed model of the user's tastes and preferences through preference elicitation. This may be done either explicitly (by querying the user) or implicitly (by observing the user's behaviour). In both cases, the cold start problem would imply that the user has to dedicate

an amount of effort using the system in its 'dumb' state - contributing to the construction of their user profile - before the system can start providing any intelligent recommendations. Content-based filtering recommender systems use information about items or users to make recommendations, rather than user preferences, so it will perform well with little user preference data. Item-based and user-based collaborative filtering makes predictions based on users' preferences for items, os they will typically perform poorly with little user preference data. Logistic regression is not recommender system technique.

NEW QUESTION: 20
The method based on principal component analysis (PCA) evaluates the features according to
A. The projection of the largest eigenvector of the correlation matrix on the initial dimensions
B. According to the magnitude of the components of the discriminate vector
C. The projection of the smallest eigenvector of the correlation matrix on the initial dimensions
D. None of the above
Answer: A (LEAVE A REPLY)
Explanation
Feature Selection:
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 discriminate analysis evaluates. Them according to the magnitude of the components of the discriminate vector.

NEW QUESTION: 21
Reducing the data from many features to a small number so that we can properly visualize it in two or three dimensions. It is done in_____
A. supervised learning
B. un-supervised learning
C. k-Nearest Neighbors
D. Support vector machines
Answer: B (LEAVE A REPLY)
Explanation
The opposite of supervised learning is a set of tasks known as unsupervised learning. In unsupervised learning, there's no label or target value given for the data. A task where we group similar items together is known as clustering. In unsupervised learning, we may also want to find statistical values that describe the data. This is known as density estimation. Another task of unsupervised learning may be reducing the data from many features to a small number so that we can properly visualize it in two or three dimensions

NEW QUESTION: 22

What describes a true property of Logistic Regression method?
A. It works well with variables that affect the outcome in a discontinuous way.
B. It is robust with redundant variables and correlated variables.
C. It works well with discrete variables that have many distinct values.
D. It handles missing values well.
Answer: (SHOW ANSWER)

NEW QUESTION: 23
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
Answer: C (LEAVE A REPLY)
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.

NEW QUESTION: 24
A data scientist wants to predict the probability of death from heart disease based on three risk factors: age, gender, and blood cholesterol level. What is the most appropriate method for this project?
A. Linear regression
B. K-means clustering
C. Logistic regression
D. Apriori algorithm
Answer: C (LEAVE A REPLY)
Explanation
Logistic regression is used widely in many fields, including the medical and social sciences. For example, the Trauma and Injury Severity Score (TRISS), which is widely used to predict mortality in injured patients, was originally developed by Boyd et al. using logistic regression. Many other medical scales used to assess severity of a patient have been developed using logistic regression. Logistic regression may be used to predict whether a patient has a given disease (e.g. diabetes; coronary heart disease), based on observed characteristics of the patient (age, sex, body mass index, results of various blood tests, etc.; age, blood cholesterol level, systolic blood pressure, relative weight, blood hemoglobin level, smoking (at 3 levels), and abnormal electrocardiogram.).Another example might be to predict whether an American voter will vote Democratic or Republican, based on age, income, sex, race, state of residence, votes in previous elections, etc. The technique can also be used in engineering, especially for predicting the probability of failure of a given process, system or

product. It is also used in marketing applications such as prediction of a customer's propensity to purchase a product or halt a subscription, etc.[citation needed] In economics it can be used to predict the likelihood of a person's choosing to be in the labor force, and a business application would be to predict the likelihood of a homeowner defaulting on a mortgage. Conditional random fields, an extension of logistic regression to sequential data, are used in natural language processing.

NEW QUESTION: 25
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 (LEAVE A REPLY)
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.

NEW QUESTION: 26
A problem statement is given as below
Hospital records show that of patients suffering from a certain disease, 75% die of it. What is the probability that of 6 randomly selected patients, 4 will recover?
Which of the following model will you use to solve it.
A. Any of the above
B. Binomial
C. Normal
D. Poisson
Answer: (SHOW ANSWER)

NEW QUESTION: 27

Suppose that we are interested in the factors that influence whether a political candidate wins an election. The outcome (response) variable is binary (0/1); win or lose. The predictor variables of interest are the amount of money spent on the campaign, the amount of time spent campaigning negatively and whether or not the candidate is an incumbent.
Above is an example of
A. Linear Regression
B. Logistic Regression
C. Recommendation system
D. Maximum likelihood estimation
E. Hierarchical linear models
Answer: B (LEAVE A REPLY)
Explanation : Logistic regression
Pros: Computationally inexpensive, easy to implement, knowledge representation easy to interpret Cons: Prone to underfitting, may have low accuracy Works with: Numeric values, nominal values

NEW QUESTION: 28
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 (LEAVE A REPLY)
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.

NEW QUESTION: 29
Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. Which of the following will you use to calculate the probability whether it will rain on the day of Marie's wedding?
A. Naive Bayes
B. Logistic Regression
C. Random Decision Forests
D. All of the above

Answer: A (LEAVE A REPLY)

Explanation

The sample space is defined by two mutually-exclusive events - it rains or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. You should consider Bayes' theorem when the following conditions exist.

* The sample space is partitioned into a set of mutually exclusive events {A1, A2,... :An}.
* Within the sample space, there exists an event B: for which P(B) > 0.
* The analytical goal is to compute a conditional probability of the form: P(Ak B).

NEW QUESTION: 30

A data scientist is asked to implement an article recommendation feature for an on-line magazine.

The magazine does not want to use client tracking technologies such as cookies or reading history. Therefore, only the style and subject matter of the current article is available for making recommendations. All of the magazine's articles are stored in a database in a format suitable for analytics.

Which method should the data scientist try first?

A. K Means Clustering

B. Naive Bayesian

C. Logistic Regression

D. Association Rules

Answer: A (LEAVE A REPLY)

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.

NEW QUESTION: 31

You are working in a classification model for a book, written by HadoopExam Learning Resources and decided to use building a text classification model for determining whether this book is for Hadoop or Cloud computing. You have to select the proper features (feature

selection) hence, to cut down on the size of the feature space, you will use the mutual information of each word with the label of hadoop or cloud to select the 1000 best features to use as input to a Naive Bayes model. When you compare the performance of a model built with the 250 best features to a model built with the 1000 best features, you notice that the model with only 250 features performs slightly better on our test data.
What would help you choose better features for your model?
A. Include least mutual information with other selected features as a feature selection criterion
B. Include the number of times each of the words appears in the book in your model
C. Decrease the size of our training data
D. Evaluate a model that only includes the top 100 words
Answer: A (LEAVE A REPLY)
Explanation
Correlation measures the linear relationship (Pearson's correlation) or monotonic relationship (Spearman's correlation) between two variables, X and Y.
Mutual information is more general and measures the reduction of uncertainty in Y after observing X.
It is the KL distance between the joint density and the product of the individual densities. So MI can measure non-monotonic relationships and other more complicated relationships
Mutual information is a quantification of the dependency between random variables. It is sometimes contrasted with linear correlation since mutual information captures nonlinear dependence.
Features with high mutual information with the predicted value are good. However a feature may have high mutual information because it is highly correlated with another feature that has already been selected.
Choosing another feature with somewhat less mutual information with the predicted value, but low mutual information with other selected features, may be more beneficial. Hence it may help to also prefer features that are less redundant with other selected features.

NEW QUESTION: 32
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 (LEAVE A REPLY)
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.

NEW QUESTION: 33
Support vector machines (SVMs) are a set of supervised learning methods used for
A. Linear classification
B. Non-linear classification
C. Regression
Answer: A,B,C (LEAVE A REPLY)
Explanation
In machine learning, support vector machines (SVMs). also support vector networks[1]) are supervised learning models with associated learning algorithms that analyze data and recognize patterns^ used for classification and regression analysis. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel tricky implicitly mapping their inputs into high-dimensional feature spaces.

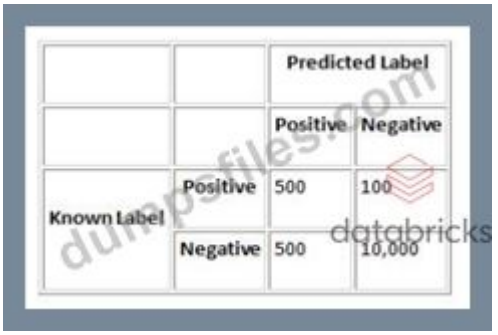
NEW QUESTION: 34
Which of the following problem you can solve using binomial distribution
A. A manufacturer of metal pistons finds that on the average: 12% of his pistons are rejected because they are either oversize or undersize. What is the probability that a batch of 10 pistons will contain no more than 2 rejects?
B. A life insurance salesman sells on the average 3 life insurance policies per week. Use Poisson's law to calculate the probability that in a given week he will sell Some policies
C. Vehicles pass through a junction on a busy road at an average rate of 300 per hour Find the probability that none passes in a given minute.

D. It was found that the mean length of 100 parts produced by a lathe was 20.05 mm with a standard deviation of 0.02 mm. Find the probability that a part selected at random would have a length between
20.03 mm and 20.08 mm
Answer: A (LEAVE A REPLY)
Explanation
The entire problem can be solved using below method
Binomial: A manufacturer of metal pistons finds that on the average, 12% of his pistons are rejected because they are either oversize or undersize. What is the probability that a batch of 10 pistons will contain no more than 2 rejects?
Poisson: A life insurance salesman sells on the average 3 life insurance policies per week. Use Poisson's law to calculate the probability that in a given week he will sell Some policies
Poisson: Vehicles pass through a junction on a busy road at an average rate of 300 per hour Find the probability that none passes in a given minute.
Normal: It was found that the mean length of 100 parts produced by a lathe was 20.05 mm with a standard deviation of 0.02 mm. Find the probability that a part selected at random would have a length between 20 03 mm and 20.08 mm

NEW QUESTION: 35
Suppose you have been given a relatively high-dimension set of independent variables and you are asked to come up with a model that predicts one of Two possible outcomes like "YES" or "NO", then which of the following technique best fit.
A. Support vector machines
B. Naive Bayes
C. Logistic regression
D. Random decision forests
E. All of the above
Answer: E (LEAVE A REPLY)
Explanation
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

NEW QUESTION: 36
Consider the following confusion matrix for a data set with 600 out of 11,100 instances positive:
In this case, Precision = 50%, Recall = 83%, Specificity = 95%, and Accuracy = 95%.

Select the correct statement



A. Precision is low, which means the classifier is predicting positives best

B. Precision is low, which means the classifier is predicting positives poorly

C. problem domain has a major impact on the measures that should be used to evaluate a classifier within it

D. 1 and 3

E. 2 and 3

Answer: E (LEAVE A REPLY)

Explanation

In this case, Precision = 50%, Recall = 83%, Specificity = 95%: and Accuracy = 95%. In this case, Precision is low, which means the classifier is predicting positives poorly. However, the three other measures seem to suggest that this is a good classifier. This just goes to show that the problem domain has a major impact on the measures that should be used to evaluate a classifier within it, and that looking at the 4 simple cases presented is not sufficient.

NEW QUESTION: 37

You are working in a data analytics company as a data scientist, you have been given a set of various types of Pizzas available across various premium food centers in a country. This data is given as numeric values like Calorie. Size, and Sale per day etc. You need to group all the pizzas with the similar properties, which of the following technique you would be using for that?

A. Association Rules

B. Naive Bayes Classifier

C. K-means Clustering

D. Linear Regression

E. Grouping

Answer: (SHOW ANSWER)

Explanation

Using K means clustering you can create group of objects based on their properties. Where K is number of the groups. In this case, in each group you determine the center of the group and then find the how far each object characteristics from the center. If it is near the center than it can be part of the group. Suppose we have 100 objects and we need to determine 4

groups. Hence, here K=4. Now we determine 4 center values and based on that center value we determine the distance of each object from the center.

NEW QUESTION: 38
Which of the following is a correct example of the target variable in regression (supervised learning)?
A. Nominal values like true, false
B. Reptile, fish, mammal, amphibian, plant, fungi
C. Infinite number of numeric values, such as 0.100, 42.001, 1000.743..
D. All of the above
Answer: D (LEAVE A REPLY)
Explanation
We address two cases of the target variable. The first case occurs when the target variable can take only nominal values: true or false; reptile, fish: mammal, amphibian, plant, fungi. The second case of classification occurs when the target variable can take an infinite number of numeric values, such as 0.100, 42.001,
1000.743, ... This case is called regression.

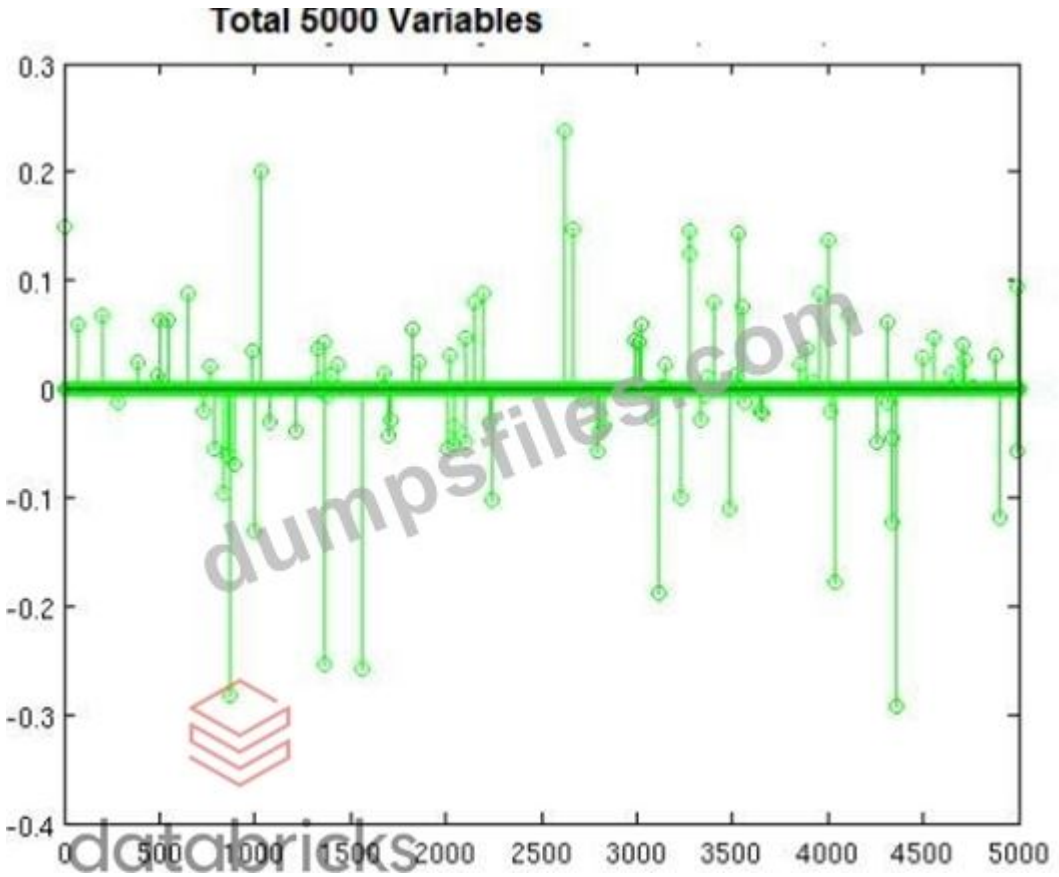
NEW QUESTION: 39
Select the statement which applies correctly to the Naive Bayes
A. Works with nominal values
B. Sensitive to how the input data is prepared
C. Works with a small amount of data
Answer: (SHOW ANSWER)

NEW QUESTION: 40
A fruit may be considered to be an apple if it is red, round, and about 3" in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the
A. Presence of the other features.
B. Absence of the other features.
C. Presence or absence of the other features
D. None of the above
Answer: (SHOW ANSWER)
Explanation
In simple terms, a naive Bayes classifier assumes that the value of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features.

NEW QUESTION: 41
RMSE is a good measure of accuracy, but only to compare forecasting errors of different models for a_____, as it is scale-dependent.
A. Between Variables
B. Particular Variable
C. Among all the variables
D. All of the above are correct
Answer: B (LEAVE A REPLY)
Explanation : The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. RMSE is a good measure of accuracy, but only to compare forecasting errors of different models for a particular variable and not between variables, as it is scale-dependent.

NEW QUESTION: 42
What is the best way to evaluate the quality of the model found by an unsupervised algorithm like k-means clustering, given metrics for the cost of the clustering (how well it fits the data) and its stability (how similar the clusters are across multiple runs over the same data)?
A. The lowest cost clustering subject to a stability constraint
B. The lowest cost clustering
C. The most stable clustering subject to a minimal cost constraint
D. The most stable clustering
Answer: A (LEAVE A REPLY)
Explanation
There is a tradeoff between cost and stability in unsupervised learning. The more tightly you fit the data, the less stable the model will be, and vice versa. The idea is to find a good balance with more weight given to the cost. Typically a good approach is to set a stability threshold and select the model that achieves the lowest cost above the stability threshold.

NEW QUESTION: 43
You are building a classifier off of a very high-dimensiona data set similar to shown in the image with 5000 variables (lots of columns, not that many rows). It can handle both dense and sparse input. Which technique is most suitable, and why?



A. Logistic regression with L1 regularization, to prevent overfitting

B. Naive Bayes, because Bayesian methods act as regularlizers

C. k-nearest neighbors, because it uses local neighborhoods to classify examples

D. Random forest because it is an ensemble method

Answer: A (LEAVE A REPLY)

Explanation

Logistic regression is widely used in machine learning for classification problems. It is well-known that regularization is required to avoid over-fitting, especially when there is a only small number of training examples, or when there are a large number of parameters to be learned. In particular L1 regularized logistic regression is often used for feature selection, and has been shown to have good generalization performance in the presence of many irrelevant features. (Ng 2004; Goodman 2004) Unregularized logistic regression is an unconstrained convex optimization problem with a continuously differentiate objective function. As a consequence, it can be solved fairly efficiently with standard convex optimization methods, such as Newton's method or conjugate gradient. However, adding the L1 regularization makes the optimization problem com-putationally more expensive to solve. If the L1 regulariza-tion is enforced by an L1 norm constraint on the parameLogistic regression is a classifier and L1 regularization tends to produce models that ignore dimensions of the input that are not predictive. This is particularly useful when the input contains many dimensions, k-nearest neighbors classification is also a classification technique, but relies on notions of distance. In a high-dimensional space, most every data point is "far" from others (the curse of dimensionality) and so these techniques break down.

Naive Bayes is not inherently regularizing. Random forests represent an ensemble method; but an ensemble method is not necessarily more suitable to high-dimensional data. Practically, I think the biggest reasons for regularization are 1) to avoid overfitting by not generating high coefficients for predictors that are sparse. 2) to stabilize the estimates especially when there's collinearity in the data.
1) is inherent in the regularization framework. Since there are two forces pulling each other in the objective function, if there's no meaningful loss reduction, the increased penalty from the regularization term wouldn't improve the overall objective function. This is a great property since a lot of noise would be automatically filtered out from the model. To give you an example for 2), if you have two predictors that have same values, if you just run a regression algorithm on it since the data matrix is singular your beta coefficients will be Inf if you try to do a straight matrix inversion. But if you add a very small regularization lambda to it, you will get stable beta coefficients with the coefficient values evenly divided between the equivalent two variables. For the difference between L1 and L2, the following graph demonstrates why people bother to have L1 since L2 has such an elegant analytical solution and is so computationally straightforward. Regularized regression can also be represented as a constrained regression problem (since they are Lagrangian equivalent). The implication of this is that the L1 regularization gives you sparse estimates. Namely, in a high dimensional space, you got mostly zeros and a small number of non-zero coefficients. This is huge since it incorporates variable selection to the modeling problem. In addition, if you have to score a large sample with your model, you can have a lot of computational savings since you don't have to compute features(predictors) whose coefficient is 0. I personally think L1 regularization is one of the most beautiful things in machine learning and convex optimization. It is indeed widely used in bioinformatics and large scale machine learning for companies like Facebook, Yahoo, Google and Microsoft.

NEW QUESTION: 44
In which of the scenario you can use the linear regression model?
A. Predicting Home Price based on the location and house area
B. Predicting demand of the goods and services based on the weather
C. Predicting tumor size reduction based on input as number of radiation treatment
D. Predicting sales of the text book based on the number of students in state
Answer: (SHOW ANSWER)
Explanation : You can use the linear regression model for predicting the continuous output variable based on the input variables. In all the cases mentioned in the question option, you can see that output can be predicted based on the input variable.
Option-A: Input: Location, House Area and Output: House Price
Option-B : Input: Weather condition, Output: Demand for the goods and services Option-C : Input: Number of Radiation Session Output: Tumor Size Reduction Option-D : Input: Number of students and Output: Sale quantity of text book

NEW QUESTION: 45
Which of the following metrics are useful in measuring the accuracy and quality of a recommender system?
A. Cluster Density
B. Support Vector Count
C. Mean Absolute Error
D. Sum of Absolute Errors
Answer: C (LEAVE A REPLY)
Explanation
The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The equation is given in the library references.
Expressed in words, the MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally in the average.
The sum of absolute errors is a valid metric, but doesn't give any useful sense of how the recommender system is performing.
Support vector count and cluster density do not apply to recommender systems.
MAE and AUC are both valid and useful metrics for measuring recommender systems.

NEW QUESTION: 46
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
Answer: A (LEAVE A REPLY)
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.

NEW QUESTION: 47
RMSE measures error of a predicted
A. For booth Numerical and categorical values
B. Numerical Value
C. Categorical values
Answer: B (LEAVE A REPLY)

NEW QUESTION: 48
Regularization is a very important technique in machine learning to prevent over fitting. And Optimizing with a L1 regularization term is harder than with an L2 regularization term because
A. The penalty term is not differentiate
B. The second derivative is not constant
C. The objective function is not convex
D. The constraints are quadratic
Answer: A (LEAVE A REPLY)
Explanation
Regularization is a very important technique in machine learning to prevent overfitting. Mathematically speaking, it adds a regularization term in order to prevent the coefficients to fit so perfectly to overfit. The difference between the L1 and L2 is just that L2 is the sum of the square of the weights, while L1 is just the sum of the weights.
Much of optimization theory has historically focused on convex loss functions because they're much easier to optimize than non-convex functions: a convex function over a bounded domain is guaranteed to have a minimum, and it's easy to find that minimum by following the gradient of the function at each point no matter where you start. For non-convex functions, on the other hand, where you start matters a great deal; if you start in a bad position and follow the gradient, you're likely to end up in a local minimum that is not necessarily equal to the global minimum.
You can think of convex functions as cereal bowls: anywhere you start in the cereal bowl, you're likely to roll down to the bottom. A non-convex function is more like a skate park: lots of ramps, dips, ups and downs. It's a lot harder to find the lowest point in a skate park than it is a cereal bowl.

NEW QUESTION: 49
A denote the event 'student is female' and let B denote the event 'student is French'. In a class of 100 students suppose 60 are French, and suppose that 10 of the French students are females. Find the probability that if I pick a French student, it will be a girl, that is, find P(A|B).
A. 1/3
B. 2/3
C. 1/6
D. 2/6
Answer: C (LEAVE A REPLY)
Explanation
Since 10 out of 100 students are both French and female, then
P(AandB)=10100
Also. 60 out of the 100 students are French, so
P(B)=60100
So the required probability is:
P(A|B)=P(AandB)P(B)=10/10060/100=16

NEW QUESTION: 50
Find out the classifier which assumes independence among all its features?
A. Neural networks
B. Linear Regression
C. Naive Bayes
D. Random forests
Answer: (SHOW ANSWER)
Explanation
A Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model".
A Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model".
In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

NEW QUESTION: 51

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: (SHOW ANSWER)
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

Valid Databricks-Certified-Professional-Data-Scientist Dumps shared by TrainingDump.com for Helping Passing Databricks-Certified-Professional-Data-Scientist Exam! TrainingDump.com now offer the **newest Databricks-Certified-Professional-Data-Scientist exam dumps**, the TrainingDump.com Databricks-Certified-Professional-Data-Scientist exam **questions have been updated** and **answers have been corrected** get the **newest** TrainingDump.com Databricks-Certified-Professional-Data-Scientist dumps with Test Engine here: https://www.trainingdump.com/Databricks/Databricks-Certified-Professional-Data-Scientist-practice-exam-dumps.html (**140** Q&As Dumps, **40%OFF Special Discount: Exam-Tests**)