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.geekcert.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

A. Naive Bayes classifier

B. Collaborative filtering

C. Logistic Regression

D. Content-based filtering

Correct Answer: B

Explanation: One scenario of collaborative filtering application is to recommend interesting or popular information as judged by the community. As a typical example, stories appear in the front page of Digg as they are "voted up" (rated positively) by the community. As the community becomes larger and more diverse, the promoted stories can better reflect the average interest of the community members.

QUESTION 2

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

Correct Answer: C

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.

QUESTION 3

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

Correct Answer: D

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.

QUESTION 4

Assume some output variable "y" is a linear combination of some independent input variables "A" plus some independent noise "e". The way the independent variables are combined is defined by a parameter vector B y=AB+e where X is an m x n matrix. B is a vector of n unknowns, and b is a vector of m values. Assuming that m is not equal to n and the columns of X are linearly independent, which expression correctly solves for B?

A. $b * (A^T * A)^{-1} * A^T$

B. $A^{-1} * b$

C. $(A^T * A)^{-1} * b$

D. $(A^T * A)^{-1} * A^T * b$

A. Option A

B. Option B

C. Option C

D. Option D

Correct Answer: D

Explanation: This is the standard solution of the normal equations for linear regression. Because A is not square, you cannot simply take its inverse.

QUESTION 5

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

Correct Answer: D

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).