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

As a data scientist consultant at ABC Corp, you are working on a recommendation engine for the learning resources for end user. So Which recommender system technique benefits most from additional user preference data?

A. Naive Bayes classifier

B. Item-based collaborative filtering

C. Logistic Regression

D. Content-based filtering

Correct Answer: B

Explanation: Item-based scales with the number of items, and user-based scales with the number of users you have. If you have something like a store, you\\'ll have a few thousand items at the most. The biggest stores at the time of writing have around 100,000 items. In the Netflix competition, there were 480,000 users and 17,700 movies. If you have a lot of users: then you\\'ll probably want to go with item-based similarity. For most product-driven recommendation engines, the number of users outnumbers the number of items. There are more people buying items than unique items for sale. Item-based collaborative filtering makes predictions based on users preferences for items. More preference data should be beneficial to this type of algorithm. Content-based filtering recommender systems use information about items or users, and not user preferences, to make recommendations. Logistic Regression, Power iteration and a Naive Bayes classifier are not recommender system techniques.

QUESTION 2

In unsupervised learning which statements correctly applies?

A. It does not have a target variable

B. Instead of telling the machine Predict Y for our data X, we\\'re asking What can you tell me about X?

C. telling the machine Predict Y for our data X

Correct Answer: AB

Explanation: In unsupervised learning we don\\'t have a target variable as we did in classification and regression.

Instead of telling the machine Predict Y for our data X, we\\'re asking What can you tell me about X?

Things we ask the machine to tell us about

X may be What are the six best groups we can make out of X? or What three features occur together most frequently in X?

QUESTION 3

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.

Correct Answer: AD

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.

QUESTION 4

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

Correct Answer: A

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:

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

QUESTION 5

You are working in an ecommerce organization, where you are designing and evaluating a recommender system, you need to select which of the following metric wilt always have the largest value?

A. Root Mean Square Error

B. Sum of Errors

C. Mean Absolute Error

D. Both land 2

E. Information is not good enough.

Correct Answer: E