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.passapply.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

Consider flipping a coin for which the probability of heads is p, where p is unknown, and our goa is to estimate p. The obvious approach is to count how many times the coin came up heads and divide by the total number of coin flips. If we flip the coin 1000 times and it comes up heads 367 times, it is very reasonable to estimate p as approximately 0.367. However, suppose we flip the coin only twice and we get heads both times. Is it reasonable to estimate p as 1.0? Intuitively, given that we only flipped the coin twice, it seems a bit rash to conclude that the coin will always come up heads, and_____is a way of avoiding such rash conclusions.

A. Naive Bayes

B. Laplace Smoothing

C. Logistic Regression

D. Linear Regression

Correct Answer: B

Explanation: Smooth the estimates: consider flipping a coin for which the probability of heads is p, where p is unknown, and our goal is to estimate p. The obvious approach is to count how many times the coin came up heads and divide by the total number of coin flips. If we flip the coin 1000 times and it comes up heads 367 times, it is very reasonable to estimate p as approximately 0.367. However, suppose we flip the coin only twice and we get heads both times. Is it reasonable to estimate p as 1.0? Intuitively, given that we only flipped the coin twice, it seems a bit rash to conclude that the coin will always come up heads, and smoothing is a way of avoiding such rash conclusions. A simple smoothing method, called Laplace smoothing (or Laplace\\'s law of succession or add-one smoothing in RandN), is to estimate p by (one plus the number of heads) / (two plus the total number of flips). Said differently, if we are keeping count of the number of heads and the number of tails, this rule is equivalent to starting each of our counts at one, rather than zero. Another advantage of Laplace smoothing is that it avoids estimating any probabilities to be zero, even for events never observed in the data. Laplace add-one smoothing now assigns too much probability to unseen words

QUESTION 2

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

Correct Answer: B

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.

QUESTION 3

You are working on a Data Science project and during the project you have been gibe a responsibility to interview all the stakeholders in the project. In which phase of the project you are?

A. Discovery

B. Data Preparations

C. Creating Models

D. Executing Models

E. Creating visuals from the outcome

F. Operationnalise the models

Correct Answer: A

Explanation: During the discovery phase you will be interviewing all the project stakeholders because they would be having quite a good amount of knowledge for the problem domain you will be working and you also interviewing project sponsors you will get to know what all are the expectations once project get completed. Hence, you will be noting down all the expectations from the project as well as you will be using their expertise in the domain.

QUESTION 4

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

Correct Answer: A

QUESTION 5

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

Correct Answer: A

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