

DS-200^{Q&As}

Data Science Essentials

Pass Cloudera DS-200 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.leads4pass.com/ds-200.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Cloudera
Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers



QUESTION 1

How can the naiveté of the naive Bayes classifier be advantageous?

- A. It does not require you to make strong assumptions about the data because it is a non- parametric
- B. It significantly reduces the size of the parameter space, thus reducing the risk of over fitting
- C. It allows you to reduce bias with no tradeoff in variance
- D. It guarantees convergence of the estimator

Correct Answer: A

QUESTION 2

You are working with a logistic regression model to predict the probability that a user will click on an ad. Your model has hundreds of features, and you're not sure if all of those features are helping your prediction. Which regularization technique should you use to prune features that aren't contributing to the model?

- A. Convex
- B. Uniform
- C. L2
- D. L1

Correct Answer: A

QUESTION 3

You have a large $m \times n$ data matrix M . You decide you want to perform dimension reduction/clustering on your data and have decide to use the singular value decomposition (SVD; also called principal components analysis PCA)

You performed singular value decomposition (SVD; also called principal components analysis or PCA) on you data matrix but you did not center your data first. What does your first singular component describe?

- A. The mean of the data set
- B. The variance of the data set
- C. The standard deviation of the data set
- D. The maximum of the data set
- E. The median of the data set

Correct Answer: C

QUESTION 4

You are about to sample a 100-dimensional unit-cube. To adequately sample any single given dimension, you need only capture 10 points. How many points do you need to order to sample the complete 100dimensional unit cube adequately?

- A. 10010
- B. 1010
- C. $\text{Log}_2(100)$
- D. 100
- E. 1000
- F. 1010

Correct Answer: E

QUESTION 5

You have a large file of N records (one per line), and want to randomly sample 10% them. You have two functions that are perfect random number generators (through they are a bit slow):

`Random_uniform ()` generates a uniformly distributed number in the interval $[0, 1]$ `random_permutation (M)` generates a random permutation of the number 0 through $M - 1$.

Below are three different functions that implement the sampling.

Method A

For line in file: `If random_uniform ()`

Method B

`i = 0`

for line in file:

`if i % 10 == 0;`

`print line`

`i += 1`

Method C

```
idxs = random_permutation (N) [: (N/10)]
```

```
i = 0
```

```
for line in file:
```

```
if i in idxs:
```

```
print line
```

```
i +=1
```

Which method might introduce unexpected correlations?

A. Method A

B. Method B

C. Method C

Correct Answer: C

[Latest DS-200 Dumps](#)

[DS-200 VCE Dumps](#)

[DS-200 Exam Questions](#)