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:

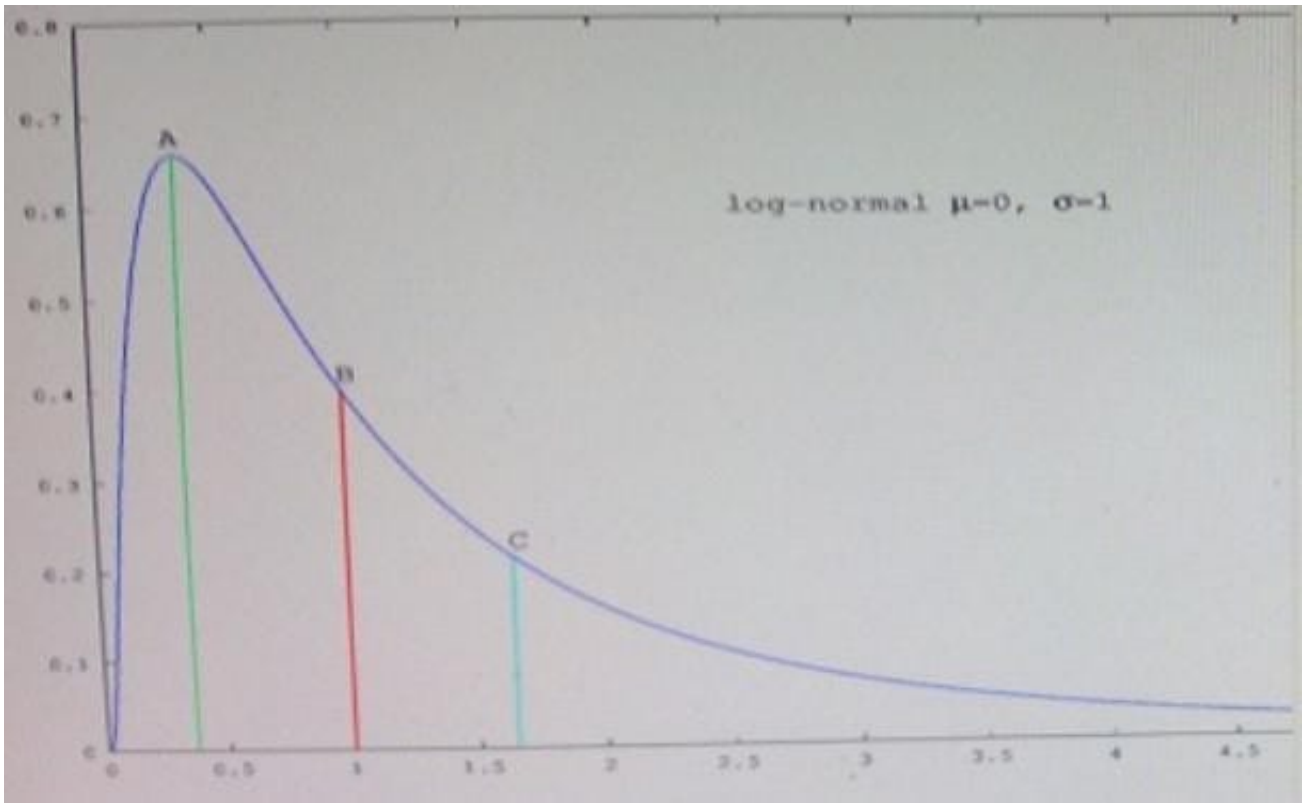
https://www.passapply.com/ds-200.html

100% Passing Guarantee
100% Money Back Assurance

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

QUESTION 1

Refer to the exhibit.



Which point in the figure is the median?

A. A

B. B

C. C

Correct Answer: A

QUESTION 2

You have a large m x 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 3

Many machine learning algorithm involve finding the Global minimum of a convex loss function, primarily because:

A. The additive inverse of a convex function is concave

B. The derivative of convex function is always defined

C. The second derivative of a convex function is a constant

D. Any local minimum of a convex is also a global minimum

Correct Answer: B

QUESTION 4

Certain individuals are more susceptible to autism if they have particular combinations of genes expressed in their DNA. Given a sample of DNA from persons who have autism and a sample of DNA from persons who do not have autism, determine the best technique for predicting whether or not a given individual is susceptible to developing autism?

A. Native Bayes

B. Linear Regression

C. Survival analysis

D. Sequence alignment

Correct Answer: B

QUESTION 5

There are 20 patients with acute lymphoblastic leukemia (ALL) and 32 patients with acute myeloid leukemia (AML), both variants of a blood cancer.

The makeup of the groups as follows:

ALL GROUP			
	Male	Female	
Caucasian	14	1	15
Asian-American	5	0	5
	19	1	20

AML GROUP			
	Male	Female	
Caucasian	9	4	13
Asian-American	7	12	19
	16	16	32

Each individual has an expression value for each of 10000 different genes. The expression value for each gene is a continuous value between -1 and 1.

You\\'ve built your model for discriminating between AML and ALL patients and you find that it works quite well on your current data. One month later, a collaboration tells you she has fresh data from 100 new AML/ ALL patients. You run the samples through your model, and turns out your model has very poor predictive accuracy on the new samples; specifically, your model predicts that all males have ALL. What is the most reliable way to fix this problem?

A. Change the distance metric

B. Reduce the number of dimensions

C. Use a Gibbs sampler on a Bayesian network

D. Perform matched sampling across other provided variables

Correct Answer: D