

100% Money Back
Guarantee

Vendor: Cloudera

Exam Code: DS-200

Exam Name: Data Science Essentials

Version: Demo

QUESTION 1

Why should stop an interactive machine learning algorithm as soon as the performance of the model on a test set stops improving?

- A. To avoid the need for cross-validating the model
- B. To prevent overfitting
- C. To increase the VC (VAPNIK-Chervonenkis) dimension for the model
- D. To keep the number of terms in the model as possible
- E. To maintain the highest VC (Vapnik-Chervonenkis) dimension for the model

Correct Answer: B

Explanation

Explanation/Reference:

QUESTION 2

What is default delimiter for Hive tables?

- A. ^A (Control-A)
- B. , (comma)
- C. \t (tab)
- D. : (colon)

Correct Answer: A

Explanation

Explanation/Reference:

Reference: <http://blog.spryinc.com/2013/10/four-useful-tricks-for-working-with-hive.html> (change the delimiter when exporting hive table)

QUESTION 3

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

Explanation

Explanation/Reference:

QUESTION 4

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

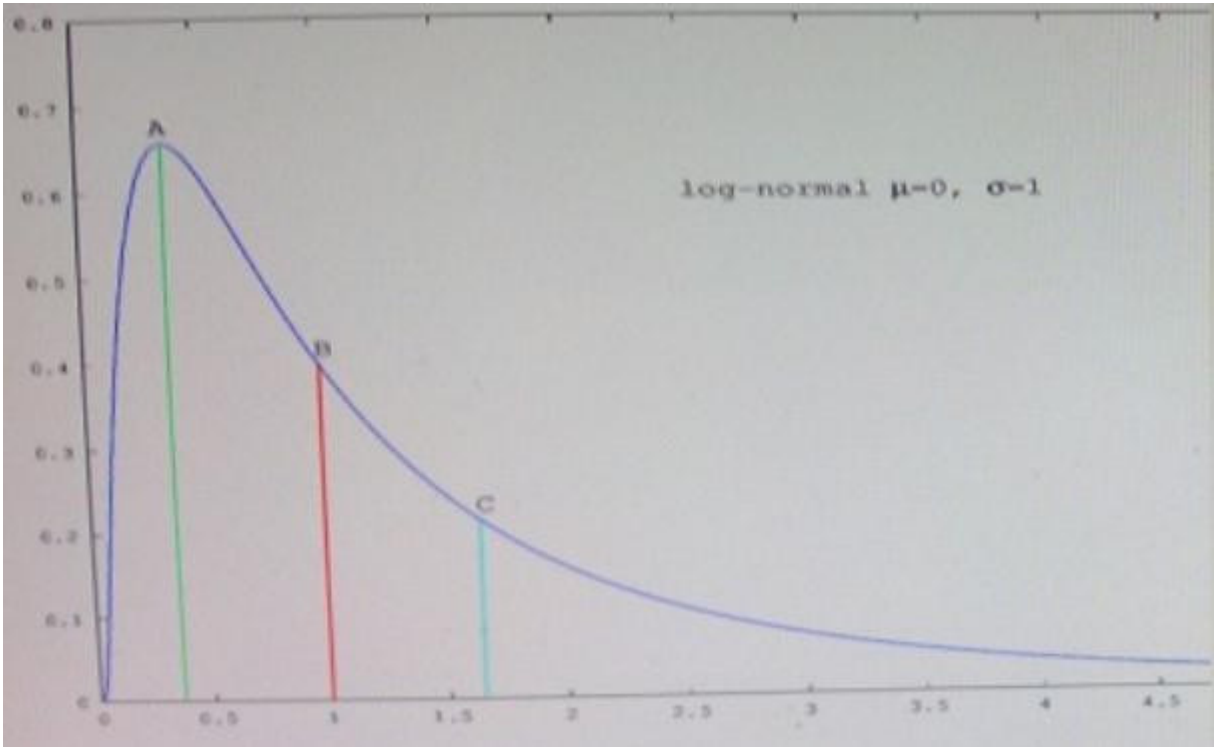
D. L1

Correct Answer: A
Explanation

Explanation/Reference:

QUESTION 5

Refer to the exhibit.



Which point in the figure is the median?

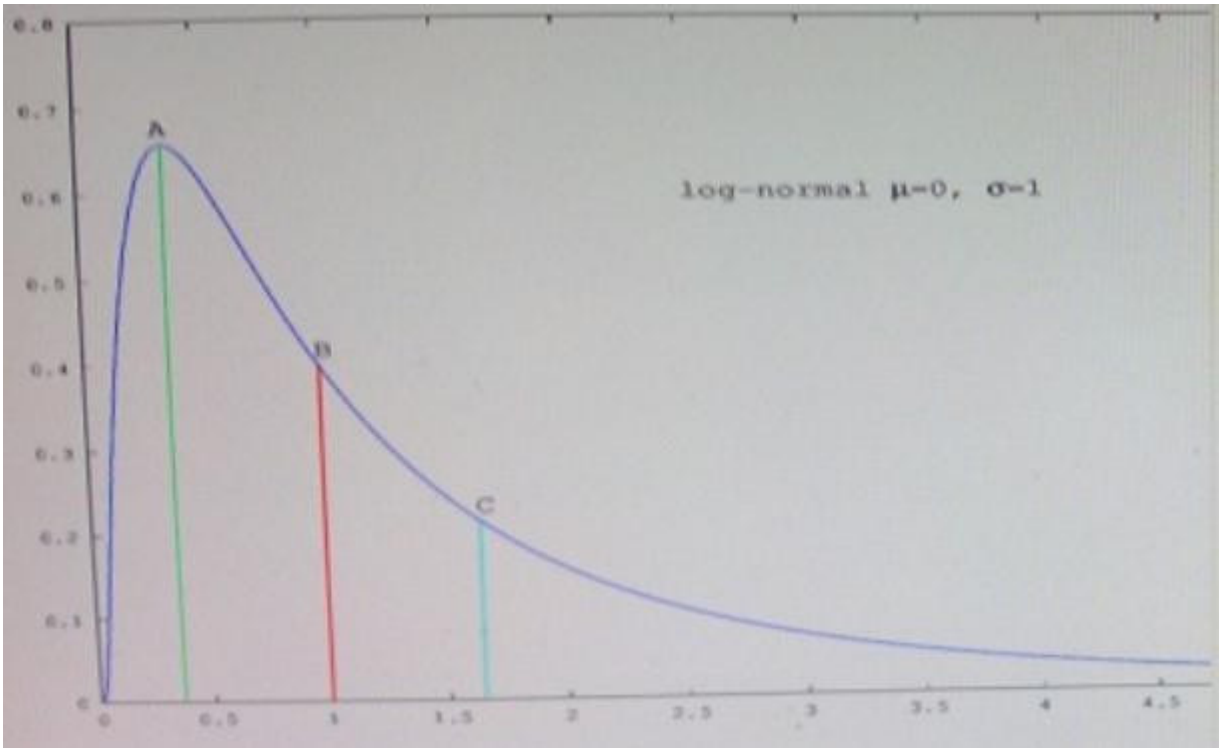
- A. A
- B. B
- C. C

Correct Answer: A
Explanation

Explanation/Reference:

QUESTION 6

Refer to the exhibit.



Which point in the figure is the mode?

- A. A
- B. B
- C. C

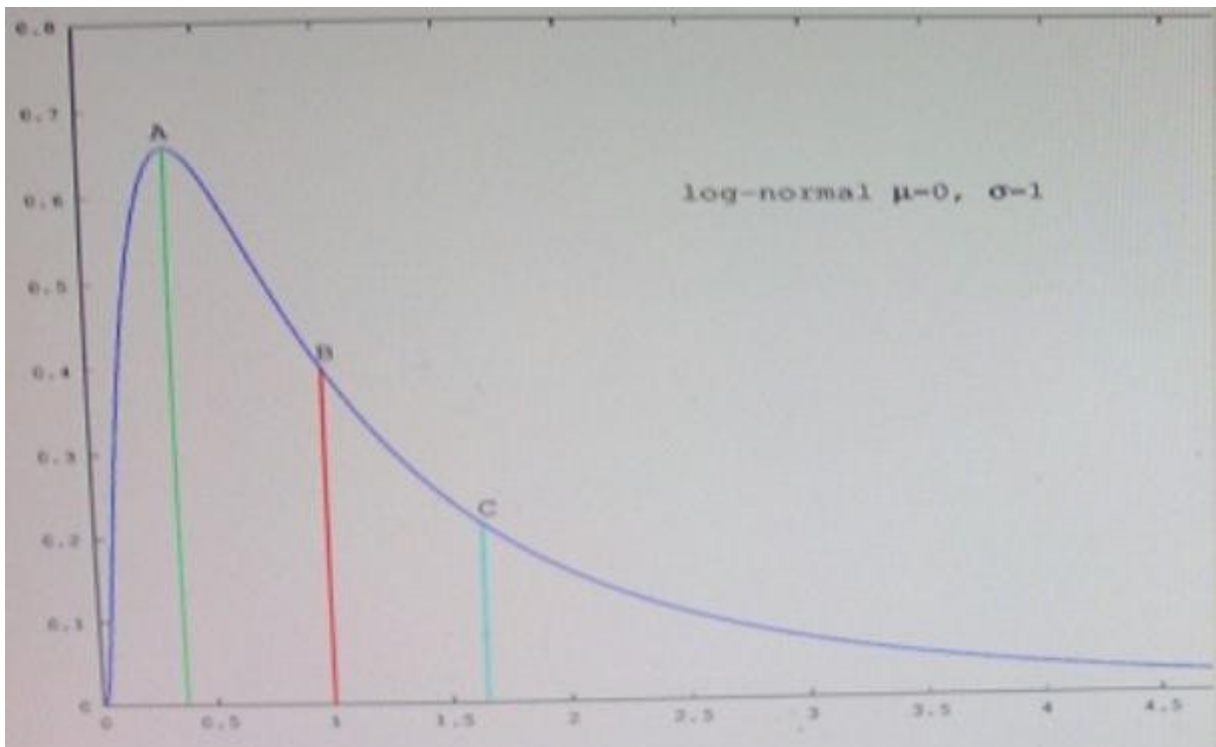
Correct Answer: C

Explanation

Explanation/Reference:

QUESTION 7

Refer to the exhibit.



Which point in the figure is the mean?

- A. A
- B. B
- C. C

Correct Answer: B

Explanation

Explanation/Reference:

QUESTION 8

Under what two conditions does stochastic gradient descent outperform 2nd-order optimization techniques such as iteratively reweighted least squares?

- A. When the volume of input data is so large and diverse that a 2nd-order optimization technique can be fit to a sample of the data
- B. When the model's estimates must be updated in real-time in order to account for new observations.
- C. When the input data can easily fit into memory on a single machine, but we want to calculate confidence intervals for all of the parameters in the model.
- D. When we are required to find the parameters that return the optimal value of the objective function.

Correct Answer: AB

Explanation

Explanation/Reference:

QUESTION 9

What is the result of the following command (the database username is foo and password is bar)?

```
$ sqoop list-tables --connect jdbc:mysql://localhost/databasename --table --username foo --password bar
```

- A. sqoop lists only those tables in the specified MySql database that have not already been imported into FDFS
- B. sqoop returns an error
- C. sqoop lists the available tables from the database
- D. sqoop imports all the tables from SQL HDFS

Correct Answer: C
Explanation

Explanation/Reference:

Reference: <https://www.inkling.com/read/hadoop-definitive-guide-tom-white-3rd/chapter-15/getting-sqoop>

QUESTION 10

What is the most common reason for a k-means clustering algorithm to return a sub-optimal clustering of its input?

- A. Non-negative values for the distance function
- B. Input data set is too large
- C. Non-normal distribution of the input data
- D. Poor selection of the initial controls

Correct Answer: C
Explanation

Explanation/Reference:

QUESTION 11

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

To Read the [Whole Q&As](#), please purchase the [Complete Version](#) from [Our website](#).

Trying our product !

- ★ **100%** Guaranteed Success
- ★ **100%** Money Back Guarantee
- ★ **365 Days** Free Update
- ★ **Instant Download** After Purchase
- ★ **24x7** Customer Support
- ★ Average **99.9%** Success Rate
- ★ More than **69,000** Satisfied Customers Worldwide
- ★ Multi-Platform capabilities - **Windows, Mac, Android, iPhone, iPod, iPad, Kindle**

Need Help

Please provide as much detail as possible so we can best assist you.

To update a previously submitted ticket:



 One Year Free Update <p>Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.</p>	 Money Back Guarantee <p>To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 Security & Privacy <p>We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information & peace of mind.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

[Guarantee & Policy](#) | [Privacy & Policy](#) | [Terms & Conditions](#)

Any charges made through this site will appear as Global Simulators Limited.

All trademarks are the property of their respective owners.

Copyright © 2004-2015, All Rights Reserved.