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.pass2lead.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

Assuming the trends shown in this chart continue, what would we expect the value of the revenue to be in Q1 of 2013?

A. $125,000

B. $170,000

C. $220,000

D. $250,000

Correct Answer: A

QUESTION 2

You are building a k-nearest neighbor classifier (k-NN) on a labeled set of points in a high- dimensional space. You determine that the classifier has a large error on the training data. What is the most likely problem?

A. High-dimensional spaces effectively make local neighborhoods global

B. k-NN compotation does not coverage in high dimensions

C. k was too small

D. The VC-dimension of a k-NN classifier is too high

Correct Answer: B

QUESTION 3

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

QUESTION 4

Which three metrics are useful in measuring the accuracy and quality of a recommender system?

A. Mutual Information

B. RMSF

C. Tanimoto coefficient

D. Pearson correlation

E. Precision

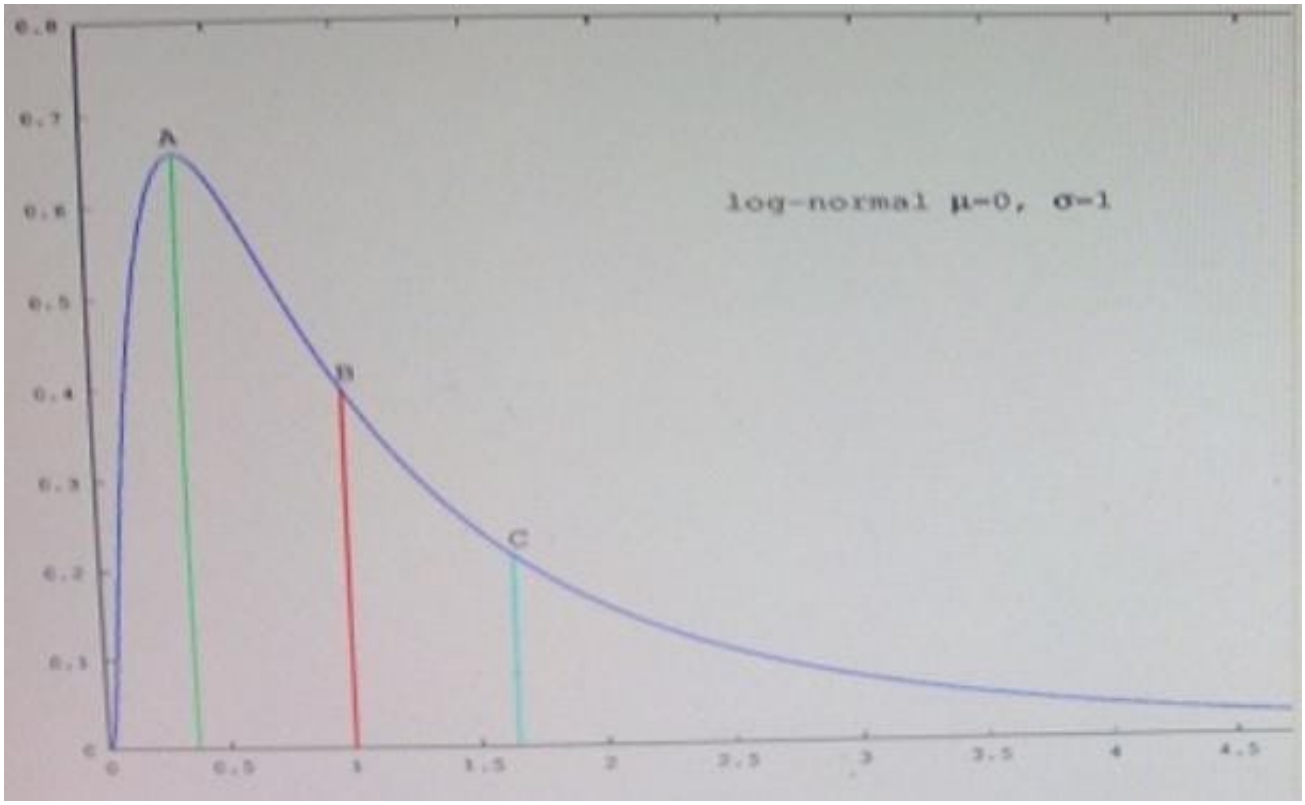
F. Recall

Correct Answer: CDE

Reference: https://lirias.kuleuven.be/bitstream/123456789/289803/3/datasets-cameraready.pdf

QUESTION 5

Refer to the exhibit.



Which point in the figure is the median?

A. A

B. B

C. C

Correct Answer: A

QUESTION 6

You want to build a classification model to identify spam comments on a blog. You decide to use the words in the comment text as inputs to your model. Which criteria should you use when deciding which words to use as features in order to contribute to making the correct classification decision?

A. Choose words for your sample that are most correlated with the Spam label

B. Choose words for your sample that occur most frequently in the text

C. Choose words, for your sample that have the largest mutual information with the spam label

D. Choose words for your sample that are least correlated with the spam label

Correct Answer: A

QUESTION 7

Which two techniques should you use to avoid overfitting a classification model to a data set?

A. Include a small number "noise" features that are not through to be correlated with the dependent variable.

B. Replicate features that are through to be significant predicators of the dependent variable multiple time for each observation.

C. Separate your input data into a training set that is used for fitting and a test set that is used for evaluating the model\\'s performance

D. Include a regularization term in the model\\'s objective function to control how precisely the model fits the data

E. Preprocess the data to exclude a typical observation from the model input

Correct Answer: AE

QUESTION 8

In what format are web server log files usually generated and how must you transform them in order to make them usable for analysis in Hadoop?

A. XML files that you need to convert to JSON

B. Text files that require parsing into useful fields

C. CSV files that require parsing into useful fields

D. HTML files that you need to convert to plain text or CSV

E. Binary files that may require decompression and conversion using AVRO

Correct Answer: AB

QUESTION 9

Why is the naive Bayes classifier "naive"?

A. It generally performs worse than more complex methods

B. It Is an unbiased estimator

C. It assumes Independence between all features

D. It makes no assumptions on the underlying distributions (i.e., it is non-parametric)

Correct Answer: C

Reference: http://www.mathworks.com/help/stats/naive-bayes-classification.html

QUESTION 10

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

Latest DS-200 Dumps DS-200 PDF Dumps DS-200 VCE Dumps