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.geekcert.com/ds-200.html

100% Passing Guarantee
100% Money Back Assurance

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

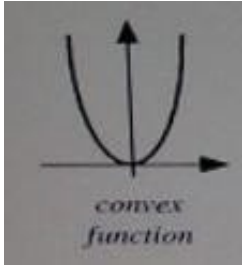⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

SATISFACTION GUARANTEED
100%
SATISFACTION GUARANTEED

QUESTION 1

Function is convex if the line segment between two points, a and b is greater than equal to the value of the a x b Which two functions are convex?

A. X1/2

B. Ex

C. 2x-1

D. 1-x2

Correct Answer: A

QUESTION 2

From historical data, you know that 50% of students who take Cloudera\\'s Introduction to Data Science: Building Recommenders Systems training course pass this exam, while only 25% of students who did not take the training course pass this exam. You also know that 50% of this exam\\'s candidates also take Cloudera\\'s Introduction to Data Science: Building Recommendations Systems training course.

What is the probability that any individual exam candidate will pass the data science exam?

A. 3/8

B. 1/4

C. 1/8

D. 1/2

Correct Answer: C

QUESTION 3

When optimizing a function using stochastic gradient descent, how frequently should you update your estimate of the

gradient?

A. Once after every pass through the data set

B. Once per observation

C. For each observation with a probability that you choose ahead of time

D. After a random number of observations

E. Once every N observations, where you decide N ahead of time

Correct Answer: AC

QUESTION 4

Which two machine learning algorithm should you consider as likely to benefit from discretizing continuous features?

A. Support vector machine

B. Naïve Bayes

C. Decision trees

D. Logistic regression

E. Singular value decomposition

Correct Answer: AB

Reference: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656082/

QUESTION 5

In what way can Hadoop be used to improve the performance of Lloyd\'s algorithm for k-means clustering on large data sets?

A. Parallelizing the centroid computations to improve numerical stability

B. Distributing the updates of the cluster centroids

C. Reducing the number of iterations required for the centroids to converge

D. Mapping the input data into a non-Euclidean metric space

Correct Answer: B

Latest DS-200 Dumps   DS-200 VCE Dumps   DS-200 Practice Test