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.pass4itsure.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

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 2

You have a directory containing a number of comma-separated files. Each file has three columns and each filename has a .csv extension. You want to have a single tab-separated file (all .tsv) that contains all the rows from all the files.

Which command is guaranteed to produce the desired output if you have more than 20,000 files to process?

A. Find . name `*, CSV\\' print0 | sargs -0 cat | tr `,\\' `\t\\' > all.tsv

B. Find . name `name * .CSV\\' | cat | awk `BEGIN {FS = "," OFS = "\t"} {print $1, $2, $3}\\' > all.tsv

C. Find . name `*.CSV\\' | tr `,\\' `\t\\' | cat > all.tsv

D. Find . name `*.CSV\\' | cat > all.tsv

E. Cat *.CSV > all.tsv

Correct Answer: B

QUESTION 3

What is the best way to determine the learning rate parameters for stochastic gradient descent when the distribution of the input data shifts over time?

A. The learning rate should be adjusted periodically based on the setting that optimizes the objective function over a sample of recent observations

B. The learning rate should be fixed number that decays as the number of observations in the data set increases

C. The learning rate should be the value that optimizes the value of the objective function over the first N samples in the dataset

D. The learning rate should be a fixed number with a constant decay factor

E. The learning rate should be continuously adjusted based on the value that optimizes the objective function for the most recent observation from the input data

Correct Answer: C

QUESTION 4

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 5

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

DS-200 PDF Dumps   **DS-200 Study Guide**   **DS-200 Braindumps**