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

QUESTION 1

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 2

You have a data file that contains two trillion records, one record per line (comma separated). Each record lists two friends and unique message sent between them. Their names will not have commas.

Michael, John, Pabst, Blue Ribbon Tiffany, James, BMX Racing John, Michael, Natural Lemon Flavor

Analyze the pseudo code examples below and determine which set of mappers and reducers in the below pseudo code snippets will solve for the mean number of messages each user sends to all of the friends?

For example pseudo code may have three friends to whom he sends 6, 10, and 200 messages, respectively, so Michael\\'s mean would be (6+10+200)/3. The solution may require a pipeline of two MapReduce jobs.

A. def mapper1 (line): key1, key2, message = line.split (` , \\') emit ((key1, key2) , 1) def reducer1(key, values): emit (key, sum(values)) def mapper2(key, value): key1, key2 = key / / unpack both friends name into separate keys emit (key1, value)

def reducer2(key, values):

emit (key, mean (values))

B. def mapper1 (line): key1, key2, message = line.split (` , \\') emit ((key1, key2) , 1) emit ((key1, key2) , 1) def reducer1(key, values): emit (key, sum(values)) def mapper2(key, value): key1, key2 = key / / unpack both friends name into separate keys emit (key1, value) def reducer2(key, values): emit (key, mean (values))

C. def mapper1 (line): key1, key2, message = line.split (` , \\') emit ((key1, key2) , 1) emit ((key1, key2) , 1) def reducer1(key, values): emit (key, sum(values))

D. def mapper (line) : Key1, key2, message = line.split (` , \\') Sort (key1, key2) / / a fiven pair will always be sorted the same Emit ((key 1, key2), 1) Def reducer1(key, values) : Emit (key, sum (values)) Def Mapper2 (key, value) Key1, key2 = key / / unpack both friends names into separate keys Emit (key1, value) Emit (key2, value) Def reducer2(key,

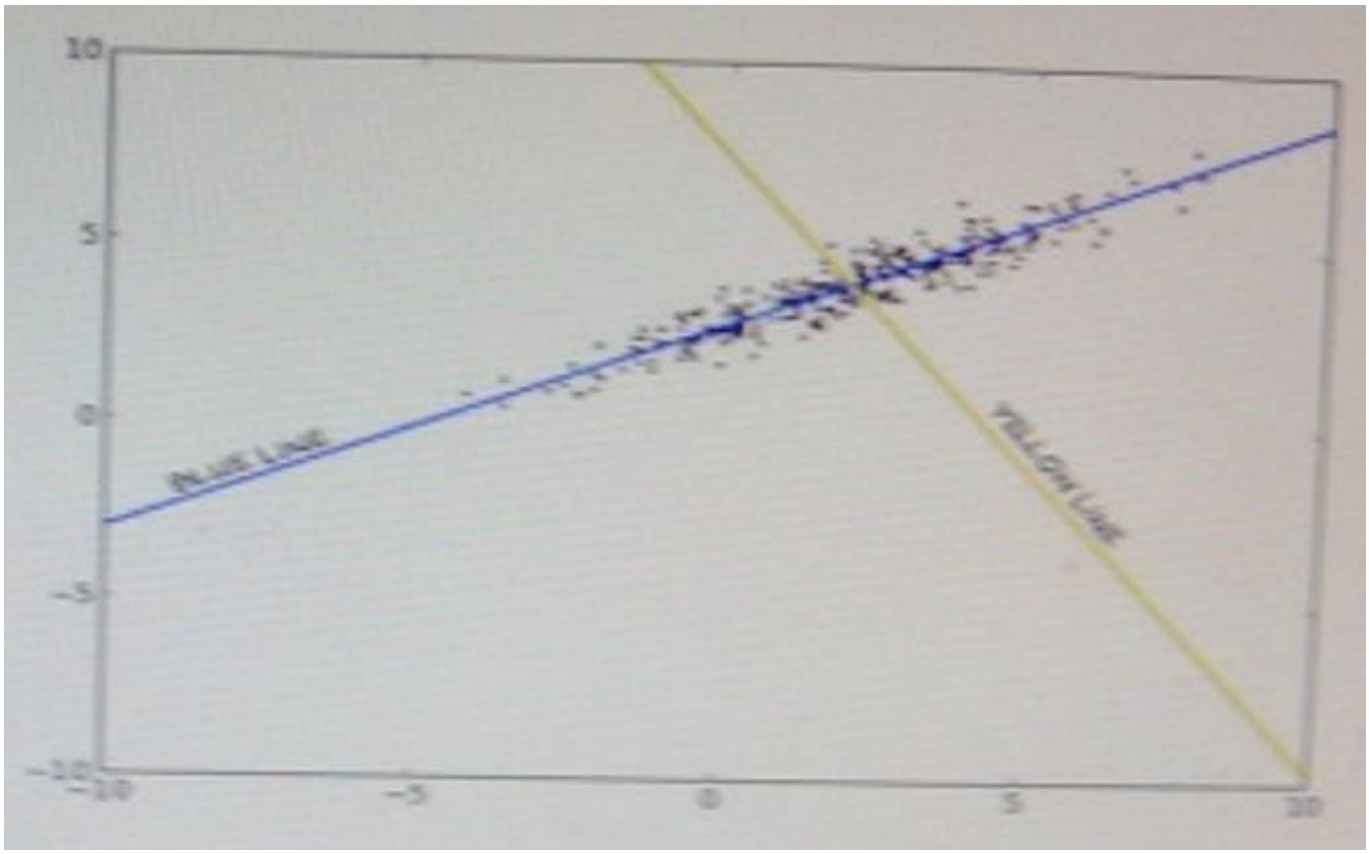
values); Emit (key, mean (values))

Correct Answer: B

QUESTION 3

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)

For the moment, assume that your data matrix M is 500 x 2. The figure below shows a plot of the data.



Which line represents the second principal component?

A. Blue

B. Yellow

Correct Answer: A

QUESTION 4

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

QUESTION 5

What is one limitation encountered by all systems that employ collaborative filtering and use preferences as input. In order to output product recommendations to consumers?

A. Consumers do not have stable ratings for the same product over time

B. There are too many consumers and too few products

C. Not every product has been rated by every consumer

D. There are too few consumers and too many products

Correct Answer: A