DC DEVELOPER
CERTIFICATION

SUCCEED AT THE
FIRST ATTEMPT!

https://developercertification.com

Questions and Answers Demo PDF

Microsoft

Exam 70-773

Analyzing Big Data with Microsoft R

Verson: Demo

[Total Questions: 10]

Question No : 1

You plan to read data from an Oracle database table and to store the data in the file system for later processing by dplyrXdf. The size of the data is larger than the memory on the server to be used for modelling.

You need to ensure that the data can be processed by dplyrXdf in the least amount of time possible.

How should you transfer the data from the Oracle database?

A. Define a data source to the Oracle database server by using RxOdbcData. Use rxImport to save the data to a comma-separated values (CSV) file.
B. Use the RODBC library, connect to the Oracle database server by using odbcConnect, and then use rxDataStep to export the data to a comma-separated values (CSV) file.
C. Define a data source to the Oracle database server by using RxOdbcData, and then use rxImport to save the data to an XDF file.
D. Use the RODBC library, connect to the Oracle database server by using odbcConnect, and then use rxSplit to save the data to multiple comma-separated values (CSV) file.

Answer: C

Question No : 2

You need to use the ScaleR distributed processing in an Apache Hadoop environment.

Which data source should you use?

A. Microsoft SQL Server database
B. XDF data files
C. ODBC data
D. Teradata database

Answer: B
Explanation:
References: https://docs.microsoft.com/en-us/machine-learning-server/r/how-to-revoscaler-hadoop

Question No : 3

You need to build a model that looks at the probability of an outcome. You must regulate between L1 and L2.

Which classification method should you use?

A. Two-Class Neutral Network
B. Two-Class Support Vector Machine
C. Two-Class Decision Forest
D. Two-Class Logistic Regression

Answer: D

Explanation:
References: https://msdn.microsoft.com/en-us/library/azure/dn905994.aspx

Question No : 4

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You build a model that uses xyz regression.

You need to estimate a model that predicts a binary variable.

Which function should you use?

A. rxPredict
B. rxLogit

C. summary
D. rxLinMod
E. rxTweedie
F. stepAic
G. rxTransform
H. rxDataStep

Answer: B

Explanation:

References: https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-logistic-regression

---

Question No : 5

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series. Information and details provided in a question apply only to that question.

You need to evaluate the significance of coefficients that are produced by using a model that was estimated already.

Which function should you use?

A. rxPredict
B. rxLogit
C. summary
D. rxLinMod
E. rxTweedie
F. stepAic
G. rxTransform
H. rxDataStep

Answer: D

Explanation:

References: https://docs.microsoft.com/en-us/machine-learning-server/r/how-to-revoscaler-

linear-model

Question No : 6

You are running a parallel function that uses the following R code segment. (Line numbers are included for reference only.)

01 cp <- 0.01 xval <- 0 maxdepth <- 5

02

03 (form, data = "segmentationDataBig", maxDepth = maxdepth, cp = cp, xval = xval, blocksPerRead = 250

You need to complete the R code. The solution must support chunking.

Which function should insert at line 02?

A. rxBTrees
B. rxExec
C. rxDForest
D. rxDTree

Answer: D

Question No : 7

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You use dplyrXdf, and you discover that after you exit the session, the output files that were created were deleted.

You need to prevent the files from being deleted.

Solution: You use rxSetComputeContext with the local parameter before performing operations that save results.

Does this meet the goal?

A. Yes
B. No

Answer: B

Question No : 8

You plan to analyze data on a local computer. To improve performance, you plan to alternate the operation between a Microsoft SQL Server and the local computer.

You need to run complex code on the SQL Server, and then revert to the local compute context.

Which R code segment should you use?

A. sqlCompute <- RxInSqlServer(connectionString = "Driver=SQL Server;Server = myServer; Database = TestDB; Uid = myID; Pwd = myPwd;")sqlPackagePaths <- RxFindPackage(package = "RevoScaleR", computeContext = sqlServerCompute)
B. sqlCompute <- RxInSqlServer(connectionstring = sqlConnString, shareDir = sqlShareDir,wait = sqlWait, consoleOutput = sqlConsoleOutput)rxSetComputeContext("local")x <- 1:10rxExec(print, x, elemType = "cores", timesToRun = 10)rxSetComputeContext("RxLocalParallel")
C. sqlCompute <- RxInSqlServer(connectionstring = sqlConnString, shareDir = sqlShareDir,wait = sqlWait, consoleOutput = sqlConsoleOutput)rxSetComputeContext("sqlCompute")x <- 1:10rxExec(print, x, elemType = "cores", timesToRun = 10)rxSetComputeContext("local")
D. sqlCompute <- RxInSqlServer(connectionstring = sqlConnString, shareDir = sqlShareDir,wait = sqlWait, consoleOutput = sqlConsoleOutput)rxSetComputeContext("local")x <- 1:10rxExec(print, x, elemType = "cores", timesToRun = 10)rxSetComputeContext("sqlCompute")

Answer: D

Explanation:

References: https://docs.microsoft.com/en-us/sql/advanced-analytics/tutorials/deepdive-define-and-use-compute-contexts

Question No : 9

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a Microsoft SQL Server instance that has R Services (In-Database) installed.

You need to monitor the R jobs that are sent to SQL Server.

Solution: You call a function from the RevoPemaR package.

Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:

References: https://docs.microsoft.com/en-us/machine-learning-server/r-reference/revopemar/pemar

Question No : 10

You are running a large logistic regression for 1,000 feature variables by using the LoisticRegression() function in the MicrosoftML package. All of the predictor variables are numeric.

Currently, you specify the input variables separately by using the following formula.

Outcome ~ Feature000 + Feature001 + Feature002 + … + Feature999

You discover that it takes 20 minutes to estimate each model.

You need to reduce the amount of time required to estimate each model without losing any information in the predictors.

What should you do?

A. Use stepControl() to perform stepwise regression to limit the number of variables that contribute to the model.
B. Use selectFeatures() to select the features that provide the most information about the outcome variable.
C. Use princomp() on the correlation matrix of Features, and then use only the first 100 principle components to reduce the number of input variables.
D. Use concat() to create a single array variable named Features, and then specify a new formula named Outcome ~ Features.

Answer: D