



Microsoft

Exam Questions 70-773

Analyzing Big Data with Microsoft R (beta)

NEW QUESTION 1

You perform an analysis that produces the decision tree shown in the exhibit.



How many leaf nodes are there on the tree?

- A. 2
- B. 3
- C. 5
- D. 7

Answer: B

NEW QUESTION 2

Note: This question is part of a series of questions that use the same or similar answer choice. 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 generate a residual based on two columns. The solution must build a trend indicator.

Which function should you use?

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

Answer: C

NEW QUESTION 3

You have following regression forest.

```
rxDForest(formula = stack.loss ~ Air.Flow + Water.Temp + Acid.Conc., data = stackloss, maxDepth = 3, nTree = 200, nTry = 2)

Type of decision forest: anova
Number of trees: 200
No. of variables tried at each split: 2
Mean of squared residuals: 44.54992 % Var explained: 65
```

Which variable contributes the most to the dependent variable?

- A. stack.loss
- B. Water.Temp
- C. Air.Flow
- D. Acid.Conc

Answer: A

NEW QUESTION 4

HOTSPOT

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You are developing a Microsoft R Open solution that will leverage the computing power of the database server for some of your datasets.

You are performing feature engineering and data preparation for the datasets. The following is a sample of the dataset.

```
rxGetInfo(df)
head(df)
```

	age	incwage	perwt	wkswork1	state
1	50	9000	30	48	Indiana
2	41	35000	20	48	Indiana
3	55	40400	21	49	Indiana
4	56	45000	30	52	Indiana
5	46	17200	60	51	Indiana
6	49	35000	21	52	Indiana

End of repeated scenario

You need to sort the data from the dataset sample and to remove duplicates by using wkswork1.

Which R code segment should you use? to answer, select the appropriate options in the answer area.

Note: Each correct selection is worth one point.

```
rxSort(inData = [sampleInData], outFile = [sampleOutData], sortByVars = c("incwage", "perwt", "age", "wkswork1"),
```

removeDupKeys = FALSE
removeDupKeys = TRUE
rxMerge = FALSE
rxMerge = TRUE

dupFreqVar
varsToDrop
VarsToKeep

```
 = "wkswork1")
```

Answer:

Explanation:

```
rxSort(inData = [sampleInData], outFile = [sampleOutData], sortByVars = c("incwage", "perwt", "age", "wkswork1"),
```

removeDupKeys = FALSE
removeDupKeys = TRUE
rxMerge = FALSE
rxMerge = TRUE

dupFreqVar
varsToDrop
VarsToKeep

```
 = "wkswork1")
```

NEW QUESTION 5

You have an Apache Hadoop Hive data warehouse. RevoScaleR is not installed. You need to sort the data according to the variables in the dataset. What should you do?

- A. Connect to the database by using an ODBC connection, and then use the rxSort function.
- B. Create a table in the ORC file format.
- C. Connect to the database by using an ODBC connection, and then use the rxDataStep function.
- D. Execute a Hive query that sorts the data, and then reads the results.

Answer: D

NEW QUESTION 6

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 section, 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

NEW QUESTION 7

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 Neural Network
- B. Two-Class Support Vector Machine

- C. Two-Class Decision Forest
- D. Two-Class Logistic Regression

Answer: A

NEW QUESTION 8

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 section, 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 dplyrXdf with the persist verb.

Does this meet the goal?

- A. Yes
- B. No

Answer: A

NEW QUESTION 9

DRAG DROP

You need to set the compute context for three different target environments.

Which Statement should you use for each environment? To answer, drag the appropriate statements to the correct execution contexts. Each statement may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Statements

RxHadoopMR()

rxSetComputeContext('local')

rxSetComputeContext('localpar')

RxSpark()

Answer Area

Parallelized execution across the cores of the edge node: server, except for rxExec calls, which are executed serially:

Parallelized execution across the cores of the edge node: server:

Parallelized distributed execution via Map Reduce across the nodes of the cluster:

Answer:

Explanation:

Statements

RxHadoopMR()

rxSetComputeContext('local')

rxSetComputeContext('localpar')

RxSpark()

Answer Area

Parallelized execution across the cores of the edge server, except for rxExec calls, which are executed s

Parallelized execution across the cores of the edge

Parallelized distributed execution via Map Reduce ac nodes of the

NEW QUESTION 10

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 section, 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 remove all instances of the file.remove method.

Does this meet the goal?

- A. Yes
- B. No

Answer: B

NEW QUESTION 10

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 section, 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 dplyrXdf with the outFile parameter and specify a path other than the working directory for dplyrXdf.

Does this meet the goal?

- A. Yes
- B. No

Answer: A

NEW QUESTION 13

HOTSPOT

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.

Start of repeated scenario

You are developing a Microsoft R Open solution that will leverage the computing power of the database server for some of your datasets.

You are performing feature engineering and data preparation for the datasets. The following is a sample of the dataset.

```
rxGetInfo(df)
head(df)
```

	age	incwage	perwt	wkswork1	state
1	50	9000	30	48	Indiana
2	41	35000	20	48	Indiana
3	55	40400	21	52	Indiana
4	56	45000	30	52	Indiana
5	46	17200	60	52	Indiana
6	49	35000	21	52	Indiana

End of repeated scenario

You plan to score some data to create data features to address empty rows. You have the following R code.

```
xdfPath <- file.path(rxGetOption("[sampleInData]", ""), "inputfile.xdf")
xdfLagged <- [sampleOutDataincludingFeatures](fileext = ".xdf")
rxSort(inData = xdfPath,
       outFile = xdfLagged,
       sortByVars = "Date")
rxDataStep(inData = xdfLagged,
           outFile = xdfLagged,
           transformObjects = list(
             varToLag = "Open",
             newName = "previousOpen"),
           transformFunc = lagVar,
           append = "cols",
           overwrite = TRUE)
rxDataStep(xdfLagged,
           varsToKeep = c("Date", "Open", "previousOpen"),
           numRows = 10)
```

You need to transform the data and overwrite the current dataset.

Which R code segment should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

The screenshot shows the configuration for the `rxDataStep` function. The `transformFunc` is set to `computeNonLagFeatures`. The `overwrite` dropdown is set to `TRUE`. The `varsToDrop` dropdown is set to `NONE`.

Answer:

Explanation:

The screenshot shows the configuration for the `rxDataStep` function. The `transformFunc` is set to `computeNonLagFeatures`. The `overwrite` dropdown is set to `TRUE`. The `varsToDrop` dropdown is set to `NONE`.

NEW QUESTION 15

DRAG DROP

You are using rxPredict for a logistic regression model.

You need to obtain prediction standard errors and confidence intervals. Which R code segment should you use?

To answer, drag the .improprate values to the correct targets. Each value may be used once, more than once, or not .You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values	Answer Area
confidence	model <- Value (default ~ year + creditScore + yearsEmploy + cDebt,
FALSE	data = trainingDataFileName, blocksPerRead = 2, verbose = 1,
glm	reportProgress=2, covCoef= Value)
none	rxPredict(model, data = targetDataFileName, outData = targetDataFileName,
rxLinMod	computeStdErr = Value , interval = " Value ", overwrite=TRUE)
rxLogit	
TRUE	

Answer:

Explanation:

Values	Answer Area
confidence	model <- rxLogit (default ~ year + creditScore + yearsEmploy + cDebt,
FALSE	data = trainingDataFileName, blocksPerRead = 2, verbose = 1,
glm	reportProgress=2, covCoef= TRUE)
none	rxPredict(model, data = targetDataFileName, outData = targetDataFileName,
rxLinMod	computeStdErr = TRUE , interval = " confidence ", overwrite=TRUE)
rxLogit	
TRUE	

NEW QUESTION 18

You need to run a larger data tree model by using rsDForest. The model must use cross validation.

Which rxDForest option should you use?

- A. maxSurrogate
- B. maxNumBins
- C. maxDepth
- D. maxCompete
- E. xVal

Answer: E

Explanation: <https://docs.microsoft.com/en-us/r-server/r/how-to-revoscaler-decision-tree>

NEW QUESTION 20

You have a dataset.

You need to repeatedly split randomly the dataset so that 80 percent of the data is used as a training set and the remaining 20 percent is used as a test set.

Which method should you use?

- A. threshold
- B. binary classification
- C. imputation
- D. cross validation
- E. pruning

Answer: D

NEW QUESTION 24

You are planning the compute contexts for your environment. You need to execute rx-function calls in parallel.

What are three possible compute contexts that you can use to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. local parallel
- B. Spark
- C. local sequential
- D. Map Reduce
- E. SQL

Answer: ABC

Explanation: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-r-server-compute-contexts>

NEW QUESTION 28

Note: This question is part of a series of questions that use the same or similar answer choice. 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 estimate a model where the outcome variable is continuous, is in the range of $[0, \infty]$, and has a substantial mass at an exact value of 0.

Which function should you use?

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

Answer: H

NEW QUESTION 32

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. Use the RODB library, connect to the Oracle database server by using odbcConnec
- B. 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 RODB library, connect to the Oracle database server by using odbcConnec
- E. and then use rxSplit to save the data to multiple comma-separated values (CSV) files.

Answer: C

NEW QUESTION 35

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 section, 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 register an Extended Events package.

Does this meet the goal?

- A. Yes
- B. No

Answer: A

NEW QUESTION 39

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

70-773 Practice Exam Features:

- * 70-773 Questions and Answers Updated Frequently
- * 70-773 Practice Questions Verified by Expert Senior Certified Staff
- * 70-773 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * 70-773 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The 70-773 Practice Test Here](#)