

PRACTICE QUESTIONS

SAS

A00-406

Machine Learning Using SAS Viya

SHARED BY [u/Sunny-Byte-5784](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	SAS A00-406, Machine Learning Using SAS Viya

Shared by u/Sunny-Byte-5784 on aiexamcerts

10 practice questions for the SAS A00-406 exam, Machine Learning Using SAS Viya, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/sas/a00-406. Spotted a mistake? Email support@aiexamcerts.com with the exam code A00-406 and the question number.

1. Which procedure is used in SAS Viya to split a CAS table into training and validation partitions with a specified proportion?
 - A PROC PRINT
 - B PROC SORT with NODUPKEY
 - C PROC PARTITION
 - D PROC APPEND

Answer: C

PROC PARTITION creates training, validation and test partitions from a CAS table, optionally stratified by a variable. PROC PRINT only displays rows, which makes it a tempting but unhelpful choice. PROC SORT with NODUPKEY removes duplicate keys, and PROC APPEND adds rows to an existing table.

2. Missing values in a numeric predictor can be imputed inside CAS, and the imputation values should be recomputed from each new scoring batch so that they reflect that batch.
 - A True
 - B False

Answer: B (False)

Imputation values estimated on the training data should be stored and applied unchanged to validation, test and production data, so the scoring pipeline matches training. Recomputing them on each batch changes the transformation for every new batch, which is the tempting but wrong practice. The statement is false.

3. Which procedure in SAS Viya provides summary statistics and distribution profiles for CAS table columns?
 - A PROC OPTMODEL
 - B PROC NNET
 - C PROC ASTORE
 - D PROC MDSUMMARY

Answer: D

PROC MDSUMMARY computes summary statistics for CAS tables, including means, minimums, maximums and counts, which supports profiling at scale. PROC OPTMODEL is an optimization procedure, and PROC NNET fits neural networks. PROC ASTORE manages stored model files, so none of them is a profiling tool.

4. The _____ of a predictor is the range of its values that are typical, and values far outside it are often flagged for review as outliers.

Answer: distribution (also accepted: interquartile range, IQR, spread)

The distribution of a predictor shows where most values lie and which values are unusually far from them. The mean is a single summary that does not show the range, which makes it the tempting alternative. The target label is the outcome being predicted rather than a description of the predictor.

5. Which two statements about neural networks are correct? (Choose two.)

- A Input variables should usually be standardized before training
- B Hidden layers allow the model to learn nonlinear combinations of inputs
- C A neural network never needs a validation partition
- D Adding layers always improves accuracy on new data
- E The number of neurons is fixed by the target variable

Answer: A, B

Standardized inputs help optimization converge, and hidden layers let the network learn nonlinear combinations of inputs. Validation remains necessary because networks can overfit, which makes the claim that no validation is needed false. Adding layers can overfit, and the number of hidden neurons is a design choice rather than a fixed value set by the target.

6. A linear model shows a large variance inflation factor for two predictors that measure almost the same thing. What is the main consequence?

- A The individual coefficients are unstable and hard to interpret
- B The model can no longer be scored on new data without retraining it first
- C The target variable is automatically rescaled
- D The training data is permanently duplicated

Answer: A

Collinearity inflates coefficient variances, so the individual effects become unstable and their signs can flip, although predictions may still be reasonable. The model remains usable for scoring, which makes the claim that the model can no longer be scored a false alarm. Target rescaling and data duplication are not consequences of collinearity.

7. Which SAS Viya procedure fits generalized linear models, such as Poisson or gamma regression, with effect selection on CAS tables?
- A PROC MIXED
 - B PROC GENSELECT
 - C PROC TIMESERIES
 - D PROC SURVEYMEANS

Answer: B

PROC GENSELECT fits generalized linear models with a chosen distribution and link and offers effect selection, running on CAS tables. PROC MIXED is a SAS 9 procedure for mixed models and is an easy mix-up, since its name also suggests modeling. PROC TIMESERIES prepares time series, and PROC SURVEYMEANS estimates means from survey data.

8. Random forests and gradient boosting models both combine many trees, but the trees in a random forest are built independently while boosting trees are built in sequence.
- A True
 - B False

Answer: A (True)

Random forests grow trees in parallel on bootstrap samples and average them, whereas boosting fits each tree to the residuals of those before it. This distinction determines their tuning and overfitting behavior. The statement is true.

9. Which SAS Viya procedure fits a gradient boosting model that builds trees sequentially, each one correcting the errors of the earlier ones?
- A PROC FOREST
 - B PROC SVMACHINE
 - C PROC ASSESS
 - D PROC GRADBOOST

Answer: D

PROC GRADBOOST fits gradient boosted trees, where each new tree fits the residual errors of the ensemble so far. PROC FOREST builds trees independently and averages them, which makes it the tempting mix-up. PROC SVMACHINE fits support vector machines and PROC ASSESS evaluates model performance.

10. Which hyperparameter controls how strongly a support vector machine penalizes misclassified training points?
- A The number of partitions used to split the training table
 - B The name of the caslib
 - C The number of output columns
 - D The cost or penalty parameter C

Answer: D

The cost parameter C sets the trade-off between a wide margin and penalties for misclassified points, so it controls how strictly the model fits the training data. The number of partitions defines the split rather than the model. The caslib name and output column count are not model settings.