

PRACTICE QUESTIONS

AWS

ME1-C02

AWS Certified Machine Learning Engineer - Associate

SHARED BY [u/Keen-Dolphin-3410](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	AWS ME1-C02, AWS Certified Machine Learning Engineer - Associate

Shared by u/Keen-Dolphin-3410 on aiexamcerts

10 practice questions for the AWS ME1-C02 exam, AWS Certified Machine Learning Engineer - Associate, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/aws/me1-c02. Spotted a mistake? Email support@aiexamcerts.com with the exam code ME1-C02 and the question number.

1. Which service delivers streaming records into Amazon S3 with buffering and optional format conversion, without managing servers?

- A Amazon Data Firehose
- B Amazon Simple Notification Service
- C AWS Batch
- D Amazon Elastic File System

Answer: A

Firehose buffers streaming records and delivers them to S3 with optional format conversion and no servers to manage. Amazon SNS is tempting because it also moves messages, but it fans out notifications rather than writing data sets. AWS Batch runs compute jobs and EFS provides shared file storage.

2. A random train-test split gives an honest estimate of accuracy for a forecasting model.

- A True
- B False

Answer: B (False)

Random splits let future records influence training, so they can overstate accuracy for forecasting. A chronological split is the honest choice. The statement is false.

3. A team must split a time-ordered dataset into training and validation sets. Which split respects the time order?

- A A random shuffle split with stratification on the date column
- B A chronological split where validation rows come after the training rows
- C Repeated random sampling with replacement from all rows
- D An alternating split that assigns every other row to validation

Answer: B

A chronological split trains on earlier records and validates on later ones, which mirrors production where the future is unknown. A random shuffle is tempting because it is the default in many libraries, but it lets the model learn from later records. Bootstrap sampling and alternating rows also mix time periods together.

4. Adding derived input columns, such as day of week computed from a timestamp, is called feature _____.

Answer: engineering

Feature engineering creates or transforms inputs so a model can learn patterns more easily. Feature selection is tempting because it also works on features, but it chooses among existing ones. Feature scaling changes magnitudes rather than creating new information.

5. Which two signals suggest a model is underfitting? (Choose two.)

- A Training error is near zero while validation error is large
- B Validation error is far below training error
- C Training and validation error are both high and close together
- D The model's predictions have very low variance
- E The model cannot reduce training error even after many epochs

Answer: C, E

Underfitting shows high error on both training and validation data, and the model fails to lower training error even with more epochs. Near-zero training error with large validation error is the overfitting pattern, which is the tempting confusion. Validation error below training error is unusual, and low prediction variance is a separate issue.

6. A regression model's residuals form a clear curved pattern when plotted against a feature. What does this most likely indicate?

- A The model is missing a nonlinear relationship
- B The training set has too many rows
- C The evaluation metric must be accuracy
- D The feature set is too small to store

Answer: A

A systematic curve in the residuals means the model's form misses a pattern, so adding nonlinear terms or a more flexible model may help. Dataset size is tempting as a cause, but size does not create structure in the errors. Metric choice and storage limits do not explain residual shape.

7. A deep model trains across several GPUs, and gradients must be synchronized after each batch. Which strategy is this?
- A Pipeline caching of prepared datasets
 - B Data parallelism
 - C Feature hashing of categorical columns
 - D Early stopping on the validation loss

Answer: B

Data parallelism replicates the model on each device, processes different batches, and synchronizes gradients. Model parallelism is tempting for very large networks, but it splits the model rather than the data. Caching, hashing, and early stopping do not synchronize gradients across devices.

8. The built-in SageMaker XGBoost algorithm requires you to write your own training container.
- A True
 - B False

Answer: B (False)

XGBoost is a built-in algorithm, so you supply data and hyperparameters without building a container. The statement is false. Custom containers are only needed for frameworks or code the built-in algorithms do not cover.

9. Which SageMaker Pipelines step evaluates a condition, such as an accuracy threshold, and routes the run down one of two branches?
- A Condition step
 - B Processing step
 - C Training step
 - D Register model step

Answer: A

A Condition step evaluates a value, such as an accuracy threshold, and decides which branch runs next. A Processing step is tempting because pipelines use it heavily, but it runs data processing code without choosing branches. Training and register steps perform their own jobs and do not route the run.

10. Which practice ensures that the model promoted to production is the one a reviewer approved?
- A Deploy whatever the latest training job produced
 - B Deploy a model version that has an Approved status in the model registry, referenced by version
 - C Copy the model file manually from a laptop to the endpoint
 - D Deploy the model from the notebook cell that trained it

Answer: B

Referencing an approved registry version ties production to the reviewed artifact and its metadata. The latest training job is tempting because it is the most recent, but it may never have been reviewed. Manual copies and notebook deployments bypass review and cannot be traced.