

PRACTICE QUESTIONS

Databricks Certified Machine Learning Associate

SHARED BY [u/placid_ibis_55](#) on aiexamcerts

EDITION Free Set 1

QUESTIONS 10

SHARED October 15, 2026

EXAM Databricks Certified Machine Learning Associate

Shared by [u/placid_ibis_55](#) on aiexamcerts

10 practice questions for the Databricks Certified Machine Learning Associate exam, each followed by its answer and a short explanation. More free sets and the full practice exam:

www.aiexamcerts.com/databricks/databricks-certified-machine-learning-associate. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. Which description best matches the machine learning lifecycle on Databricks?

- A A one-time process in which a model is trained once and never revisited.
- B A linear process in which deployment happens before any evaluation.
- C An iterative cycle from data preparation through training, evaluation, deployment and monitoring.
- D A hand-off process in which only the data engineer takes part after the first notebook is written.

Answer: C

The lifecycle runs from data preparation through training, evaluation, registration and deployment, and production models are monitored and retrained as needed. Treating training as a one-time event is the most tempting misconception, since many first models are never revisited. Deployment comes after evaluation, and machine learning work involves more roles than data engineering.

2. In MLflow, a named collection of runs that groups related training attempts is called an _____.

Answer: experiment

An experiment groups runs for one modeling question, and the UI and API let you filter and compare the runs inside it. A run is a single execution within an experiment, which is the tempting but narrower term. A registered model is a separate object in the model registry.

3. Which Databricks object stores precomputed features as a Delta table with a primary key, so that training and serving can reuse them?

- A A registered model version in Unity Catalog.
- B An MLflow experiment with logged metrics.
- C A volume in Unity Catalog that stores raw files and models.
- D A feature table in Unity Catalog with a primary key.

Answer: D

A feature table is a governed Delta table keyed by a primary key, which lets lookups retrieve the features for each entity. A registered model version stores a trained model, and an MLflow experiment groups runs and metrics. A volume holds files and does not serve keyed features.

4. Delta Lake time travel lets you query earlier versions of a table with VERSION AS OF or TIMESTAMP AS OF, but only the most recent version is ever retained.
- A True
 - B False

Answer: B (False)

Delta keeps earlier table versions in its transaction log until old files are removed by VACUUM and the retention period passes, which is what makes time travel possible. The statement's claim that only the latest version is kept is false.

5. Which two statements about bias and variance in machine learning models are correct? (Choose two.)
- A High variance usually shows up as a large gap between training and validation performance.
 - B Adding more regularization generally increases variance.
 - C High bias means the model underfits the training data.
 - D Deeper trees always reduce bias and variance together.
 - E A model with low training error must have low validation error.

Answer: A, C

A model that fits training data far better than validation data is showing high variance, while high bias appears as poor fit even on training data. Regularization usually reduces variance at some cost in bias, so the claim that more regularization raises variance has the relationship backwards. Low training error does not guarantee low validation error, which is the tempting overreach.

6. Which scikit-learn object chains preprocessing steps and a final estimator so that they are fitted together on training data only?
- A sklearn.metrics.make_scorer
 - B sklearn.datasets.load_iris
 - C sklearn.pipeline.Pipeline
 - D sklearn.decomposition.PCA used alone

Answer: C

A scikit-learn Pipeline fits each step on the training data and applies the fitted steps to new data, which prevents preprocessing leakage. PCA alone performs one transformation and does not chain with the estimator, which makes it a tempting but incomplete answer. The scorer and the dataset loader serve different purposes.

7. What is an MLflow artifact?

- A A named group of related runs in one workspace folder.
- B A single hyperparameter value logged at the start of each run.
- C An output file attached to a run, such as a model or plot.
- D A numeric metric logged at each step of model training.

Answer: C

Artifacts are files stored with a run, including serialized models, figures and data samples, and they can be downloaded later. A parameter is a single logged value, which makes it the tempting confusion. An experiment groups runs and a metric is a number logged per step, so neither is a stored file.

8. An MLflow run that stops because of an error has its status set to _____.

Answer: FAILED

A run marked FAILED has an error that ended it, while KILLED is used when a user or system deliberately stopped it. Both states show up in the run list, so filtering by status helps find problems. FINISHED is reserved for runs that completed normally.

9. Which Hyperopt search-space function draws a value from a set of discrete options?

- A hp.uniform
- B hp.loguniform
- C hp.choice
- D Trials

Answer: C

hp.choice selects one value from a list of options, which suits categorical hyperparameters such as the type of kernel. hp.uniform and hp.loguniform sample continuous ranges, which is the tempting confusion for numeric settings. Trials records results and does not define a search space.

10. A team needs a quick baseline for a classification problem with minimal code. Which option is most appropriate?

- A Building a custom deep learning architecture from scratch.
- B Databricks AutoML classification.
- C Training a model on a separate cluster without any tracking.
- D Hand-coding a distributed training loop with Horovod.

Answer: B

AutoML classification searches models and settings automatically and logs each trial, which gives a fast, traceable baseline. Building a custom architecture or a Horovod training loop is far more work than a baseline needs. Untracked training loses the record that makes a baseline useful.