

PRACTICE QUESTIONS

DASCA

Senior Data Scientist

SHARED BY [u/tidy_anchor_93](#) on aiexamcerts

EDITION	Free Set 1
QUESTIONS	10
SHARED	October 15, 2026
EXAM	DASCA Senior Data Scientist

Shared by [u/tidy_anchor_93](#) on aiexamcerts

10 practice questions for the DASCA Senior Data Scientist exam, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/dasca/senior-data-scientist. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. An A/B test team checks results every day and stops the experiment as soon as the p-value dips below 0.05. What is the main statistical risk of this practice?
 - A The sample size will be too large to analyze.
 - B Results will be biased toward the treatment group automatically, because early winners get more traffic.
 - C The false positive rate rises above the nominal alpha because of repeated looks at the data.
 - D A t-test can no longer be used on the final data.

Answer: C

Each interim look is another chance for noise to cross the threshold, so the true Type I error rate grows well past 0.05. Sequential designs with alpha spending or fixed sample sizes are the standard remedy. Sample size and test choice are not the problem, and the treatment group is not favored by peeking.

2. The probability of rejecting a true null hypothesis is the _____ level of a test.

Answer: significance (also accepted: alpha, significance level, significance level (alpha))

The significance level, usually written alpha, is the Type I error rate the analyst is willing to accept. Power is the probability of correctly rejecting a false null, which is the tempting but different quantity. Beta is the Type II error rate.

3. A gradient boosting model is trained with a very small learning rate and too few trees. What is the most likely symptom?
 - A Underfitting, because each tree contributes little.
 - B Overfitting, because the trees are allowed to grow very deep.
 - C Data leakage introduced by the learning rate.
 - D Exploding gradients during tree construction.

Answer: A

A small learning rate combined with too few rounds leaves the ensemble far from fitting the signal, which shows up as underfitting on both training and validation data. Overfitting is the tempting opposite, but it comes from too much capacity or too many rounds. Learning rate does not cause leakage, and tree construction does not suffer from exploding gradients.

4. Macro-averaged F1 weights every class equally, so it can expose poor performance on a rare class that overall accuracy hides.

- A True
- B False

Answer: A (True)

Macro averaging computes the F1 score for each class and takes the unweighted mean, so a rare class with poor F1 pulls the score down. Overall accuracy is dominated by the frequent classes. The statement is true.

5. Which two statements about reproducing neural network training runs are correct? (Choose two.)

- A Setting random seeds removes all nondeterminism from every GPU kernel.
- B Recording library versions and hardware details helps explain differences between runs.
- C The data order should be randomized differently on each rerun to improve reproducibility.
- D Logging hyperparameters and the code commit makes runs comparable.
- E A single run is sufficient to report a model's performance.

Answer: B, D

Versions, hardware and configuration determine whether a rerun can match the original, and logging the code commit and hyperparameters makes runs comparable. Seeds do not remove every source of GPU nondeterminism, which makes the claim that they do a tempting overclaim. Reporting one run hides variation across seeds.

6. Which technique adds small trainable low-rank matrices to frozen model weights to reduce the cost of fine-tuning?

- A Gradient clipping
- B Label smoothing
- C LoRA
- D Early stopping

Answer: C

LoRA, low-rank adaptation, freezes the pretrained weights and learns small low-rank update matrices, which sharply reduces trainable parameters and memory. Gradient clipping stabilizes training, label smoothing changes targets, and early stopping halts training. None of these three changes which parameters are trained.

7. Partitioning a large table by date mainly helps which query pattern?
- A Queries that join only on customer ID.
 - B Writes that require a global table lock.
 - C Queries that filter on date, which can then scan fewer files.
 - D Full table exports to a single CSV file.

Answer: C

Partition pruning lets the engine skip directories or files whose date values cannot match the filter, which reduces the data scanned. Joins on other keys do not benefit directly, which makes that answer a tempting but weaker choice. Partitioning does not create table locks, and full exports read every partition.

8. A columnar file format that supports efficient analytical scans is Apache _____.

Answer: Parquet

Apache Parquet stores data by column with compression and per-column statistics, so analytic queries read only the columns they need. Row-oriented text formats such as CSV force full row reads, which is the tempting alternative. Avro is also an Apache format but is primarily row-oriented.

9. Which condition most justifies retraining a deployed model?
- A Its input distribution or error rate has shifted since training.
 - B Its launch test accuracy was high, so it should never need retraining.
 - C Its training loss decreased during the last epoch.
 - D A newer GPU type became available.

Answer: A

Drift in inputs or degradation in outcomes means the training data no longer represents the environment, which is the main trigger for retraining. High launch accuracy is a snapshot and does not guarantee future performance, which is the tempting misreading. Training loss and hardware availability are not evidence about the model's fit to current data.

10. In MLOps, what is a model card?
- A A GPU accelerator installed in the training cluster and used for every model run in the lab.
 - B An API key file for the serving endpoint.
 - C A document describing a model's intended use, data, evaluation and limits.
 - D A dashboard widget that shows uptime.

Answer: C

A model card records the facts a reviewer needs to use a model responsibly, including its intended use, data, metrics and known limits. Its purpose is transparency rather than hardware or operations. Uptime widgets and API keys are operational artifacts, not documentation of the model.