

PRACTICE QUESTIONS

AI CERTs

AI+ Engineer

SHARED BY [u/grandlynx1999](#) on aiexamcerts

EDITION	Free Set 1
QUESTIONS	10
SHARED	October 15, 2026
EXAM	AI CERTs AI+ Engineer

Shared by [u/grandlynx1999](#) on aiexamcerts

10 practice questions for the AI CERTs AI+ Engineer exam, each followed by its answer and a short explanation. More free sets and the full practice exam: www.aiexamcerts.com/ai-certs/ai-engineer. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. Which metric is most appropriate for evaluating a regression model that predicts house prices in dollars?
- A Accuracy
 - B Root mean squared error
 - C F1 score
 - D Area under the ROC curve

Answer: B

Root mean squared error measures the typical size of prediction errors in the same units as the target, which suits a continuous price. Accuracy is the tempting choice, but it is a classification metric and needs discrete class labels. F1 score and ROC AUC also belong to classification, so they do not describe numeric price errors.

2. The adaptive optimizer that keeps running estimates of the first and second moments of each gradient is _____.

Answer: Adam (also accepted: Adam optimizer)

Adam combines momentum with per-parameter scaling from squared gradients, which is why it is a common default. Stochastic gradient descent is a simpler optimizer and does not keep these per-parameter estimates.

3. A model scores 99 percent on its training data and 62 percent on an unseen validation set. What is the most likely diagnosis?
- A Underfitting caused by too few features
 - B Data leakage from the test set
 - C Overfitting to the training data
 - D Class imbalance in the labels

Answer: C

A large gap between training and validation performance is the classic signature of overfitting, where the model memorized patterns that do not generalize. Underfitting would show poor performance on the training data as well. Data leakage would usually inflate validation scores rather than depress them.

4. Early stopping monitors validation performance to decide when to halt training.
- A True
 - B False

Answer: A (True)

This is true. Training halts when validation metrics stop improving for a set number of checks, which helps capture the model before it overfits the training data.

5. Which two actions help prevent data leakage in a preprocessing pipeline? (Choose two.)
- A Fitting imputers and scalers on the training split only
 - B Computing target-based features using the full dataset before splitting
 - C Shuffling the test set so it resembles the training set
 - D Wrapping preprocessing and model training in one pipeline that is cross-validated
 - E Using test labels to choose feature thresholds

Answer: A, D

Fitting transformers only on training folds keeps statistics from the evaluation data out of the model. A single pipeline that is cross-validated repeats that discipline inside every fold. Computing target-based features before splitting and using test labels for decisions both leak information, and shuffling the test set does not change what the model has learned.

6. In k-fold cross-validation, what is the main purpose?
- A Estimating generalization with rotating held-out folds
 - B Permanently increasing the size of the training dataset
 - C Selecting the activation function automatically
 - D Converting categorical features into numeric codes

Answer: A

Each fold takes a turn as the held-out validation set while the others train the model, and the scores are averaged for a steadier estimate. Permanently enlarging the training data is a common misreading, since the original data is only partitioned. The other two options describe activation search and categorical encoding, which cross-validation does not perform.

7. Why should a scaler such as a standard scaler be fitted only on the training split?

- A Scaling is only allowed on categorical data
- B Fitting on all data leaks test statistics
- C Scaled values cannot be stored for later use
- D The test set must always stay in its raw units

Answer: B

If the scaler learns its mean and standard deviation from the test split, information about unseen data leaks into training and inflates the evaluation. Scaling applies to numeric features, not categorical ones, which is a different misconception. The test set is transformed with the training statistics so that it is measured on the same scale as the model.

8. Splitting text into the units a model reads and generates is called _____.

Answer: tokenization

Tokenization converts raw text into tokens drawn from the model's vocabulary, and the same tokenizer must be used at training and inference time. Lemmatization is a different preprocessing step that reduces words to dictionary forms.

9. Which activation function is most commonly used in hidden layers of modern deep networks to reduce vanishing gradients compared with sigmoid?

- A Softmax
- B ReLU
- C Step function
- D Linear identity

Answer: B

ReLU passes positive inputs with a constant gradient of one, which avoids the saturation that makes sigmoid gradients shrink in deep stacks. Softmax is tempting because it appears in deep networks, but it is normally used at the output to produce class probabilities. A step function has no useful gradient, and a linear identity adds no nonlinearity.

10. Which technique lets a language model answer using documents retrieved at query time rather than only its training weights?

- A Quantization
- B Retrieval-augmented generation
- C Knowledge distillation into a smaller model
- D Beam search

Answer: B

Retrieval-augmented generation fetches relevant passages from an index and places them in the prompt so the answer can use current or private content. Quantization reduces numeric precision and does not add knowledge. Distillation trains a smaller model from a larger one, and beam search is a decoding strategy.