

PRACTICE QUESTIONS

AI CERTs

AI+ Engineer

SHARED BY [u/goldenbadger1984](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	AI CERTs AI+ Engineer

Shared by [u/goldenbadger1984](#) 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 regularization method tends to drive some coefficients exactly to zero, producing a sparse model?
- A L2 regularization (ridge)
 - B L1 regularization (lasso)
 - C Early stopping
 - D Data augmentation

Answer: B

L1 regularization adds the absolute value of the coefficients to the loss, and that penalty can push weak coefficients to exactly zero. L2 regularization is the tempting choice because it is also common, but it shrinks coefficients toward zero without usually eliminating them. Early stopping and data augmentation do not penalize coefficients at all.

2. Dropout is applied at inference time to randomly remove neurons and improve speed.
- A True
 - B False

Answer: B (False)

This is false. Dropout is a training-time regularizer, and standard inference uses all units with scaling so the output is deterministic and not sped up by removing neurons.

3. A fraud model flags very few transactions, and nearly every flagged transaction turns out to be genuine fraud. Which metric is high in this situation?
- A Recall
 - B Accuracy
 - C Precision
 - D Mean squared error

Answer: C

Precision is the share of flagged transactions that are truly fraudulent, calculated as true positives divided by all predicted positives. Recall is the tempting answer, but it measures the share of actual fraud the model caught, which is low when few cases are flagged. Accuracy and mean squared error do not isolate the quality of the flagged set.

4. Replacing each category with an integer code is called _____ encoding.

Answer: label

Label encoding maps each category to an integer and is compact but implies an order. Its tempting alternative, one-hot encoding, creates one binary column per category instead.

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

- A Adding more representative training data often reduces it
- B It is best diagnosed by low training error and low validation error together
- C It is cured by simply increasing model complexity
- D It appears as low training error paired with much higher validation error
- E It is caused only by using too few hidden layers

Answer: A, D

Overfitting shows as a model that fits training data well but generalizes poorly, and the widening gap to validation error is its usual signature. More representative data helps because the model has less room to memorize noise. Low error on both sets indicates good fit rather than overfitting, and adding complexity makes overfitting worse, not better. Layer count is only one possible cause.

6. Which encoding converts a nominal category such as color into separate binary columns, one per category?

- A Ordinal encoding
- B Label smoothing
- C One-hot encoding
- D Min-max normalization

Answer: C

One-hot encoding creates one binary column per category so that no artificial order is implied between values such as red, green and blue. Ordinal encoding is the tempting choice, but it assigns integers that suggest a ranking the category does not have. Label smoothing and min-max normalization are unrelated to converting categories into columns.

7. Which network architecture is most associated with processing grid-like data such as images through learned spatial filters?
- A Recurrent neural network
 - B Restricted Boltzmann machine
 - C Transformer decoder only
 - D Convolutional neural network

Answer: D

Convolutional neural networks slide learned filters across an image, sharing weights across positions to detect edges and textures. Recurrent networks process sequences step by step and are a tempting but mismatched choice for spatial grids. Restricted Boltzmann machines and decoder-only transformers are generative or sequence models rather than the standard image architecture.

8. Cross-validation gives a more stable performance estimate than a single train-test split.
- A True
 - B False

Answer: A (True)

This is true. Averaging results over several folds reduces the dependence of the estimate on one lucky or unlucky partition of the data.

9. What does the context window of an LLM define?
- A The maximum number of GPUs used in parallel during training
 - B The number of training examples per batch
 - C The tokens the model can attend to in one request
 - D The time limit for a generated response

Answer: C

The context window is the limit on how many tokens, including the prompt and the generated output, the model can process in a single request. Batch size is a training setting that is easily confused with it, but it has nothing to do with inference-time input length. GPU count and response time limits are deployment concerns rather than the context window.

10. Which tokenization method is commonly used by GPT-style models and builds its vocabulary by merging frequent byte pairs?
- A WordNet lexical database
 - B Stemming of word forms
 - C Lemmatization of word forms
 - D Byte-pair encoding

Answer: D

Byte-pair encoding starts from characters or bytes and repeatedly merges the most frequent adjacent pairs to build a subword vocabulary. WordNet is a lexical database rather than a tokenizer, which makes it a tempting but wrong pick. Stemming and lemmatization reduce words to base forms and do not define the model's token vocabulary.