

PRACTICE QUESTIONS

IABAC

Certified Artificial Intelligence Expert

SHARED BY [u/Merry-Wombat-8141](#) on aiexamcerts

EDITION Free Set 1

QUESTIONS 10

SHARED October 15, 2026

EXAM IABAC Certified Artificial Intelligence Expert

Shared by [u/Merry-Wombat-8141](#) on aiexamcerts

10 practice questions for the IABAC Certified Artificial Intelligence Expert exam, each followed by its answer and a short explanation. More free sets and the full practice exam:

www.aiexamcerts.com/iabac/certified-artificial-intelligence-expert. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. Why are residual connections important in very deep networks?

- A They remove the need for any non-linear activation function in the block
- B They give gradients identity paths that ease training of deep stacks
- C They guarantee that the loss function stays convex during training
- D They normalize each mini-batch to zero mean and unit variance

Answer: B

Skip connections add the input of a block to its output, giving gradients a direct path back through the network. Removing activation functions is the tempting distractor, since residual blocks look simpler, but they still rely on non-linearities. Convexity guarantees and batch normalization are not what residual connections provide.

2. The activation function that outputs the input when it is positive and zero otherwise is called the rectified linear unit, abbreviated _____.

Answer: ReLU

ReLU is a simple non-linearity that avoids saturation for positive inputs and is widely used in deep networks. Sigmoid is the tempting answer, since it is also a common activation, but it squashes inputs into the range between zero and one. Softmax normalizes a vector of scores into probabilities.

3. What does a learning rate warm-up schedule do at the start of training?

- A It sets the learning rate to its largest value immediately
- B It freezes all pretrained weights until the final epoch of the training run
- C It halves the learning rate after every epoch from the start
- D It raises the learning rate gradually from a small starting value

Answer: D

Warm-up avoids large unstable updates while the weights and optimizer statistics are still poorly estimated. Starting at the largest value is the tempting distractor, since fast early progress seems desirable, but it often destabilizes training. Freezing weights and halving the rate every epoch are not warm-up behavior.

4. A larger batch size always improves generalization, because it lowers the noise in the gradient estimates.

- A True
- B False

Answer: B (False)

Larger batches give smoother gradient estimates, yet they can converge to sharper minima that generalize less well. A bigger batch always improving test accuracy is the tempting belief. The statement is false.

5. Which two statements about cross-validation for time-ordered data are correct? (Choose two.)

- A Random shuffling can leak future information into training
- B Standard random k-fold is always the correct choice for forecasting
- C Time-ordered data never needs a validation split
- D Folds should respect time order so that training data precedes validation data
- E Future observations are a helpful feature for predicting the same period

Answer: A, D

Forward-chaining or blocked splits mirror real deployment, and random shuffling lets future patterns influence the past. Standard random k-fold is the tempting distractor, since it is the default in many tutorials, but it can overstate forecasting accuracy. Validation splits remain necessary, and using future observations is leakage.

6. What is a feature store mainly used for?

- A Managing feature definitions and serving the same values for training and inference
- B Storing the final weights of trained models together with their evaluation history and logs
- C Tracking hyperparameter search results across training runs
- D Labeling raw examples through a human review queue

Answer: A

A feature store centralizes feature computation, versioning and retrieval so teams reuse the same definitions. Model registries are the tempting distractor, since they also store artifacts, but they hold trained models rather than feature values. Experiment tracking and labeling are not its role.

7. Why is point-in-time correctness important when building a training dataset from event logs?
- A It guarantees that every event in the logs has the same number of columns
 - B It prevents features from using information not yet known at label time
 - C It balances the class counts across every training fold
 - D It makes the dataset smaller than the raw logs

Answer: B

Joining only the feature values available at each label time avoids leakage from the future. Equal column counts is the tempting distractor, since consistent schemas matter, but they say nothing about temporal leakage. Class balancing across folds is a separate concern from temporal leakage.

8. A common method for estimating model variance is to resample the training data with replacement, which is called _____.

Answer: bootstrapping (also accepted: bootstrap)

Bootstrapping draws samples with replacement to estimate the spread of a statistic or model result. Cross-validation is the tempting answer, since it also estimates performance, but it partitions data into folds rather than resampling with replacement. Permutation testing shuffles labels to test significance.

9. Which component of a transformer block is typically applied independently to each position?
- A The position-wise feed-forward network
 - B The attention softmax across all positions
 - C The causal mask matrix
 - D The tokenizer's merge table

Answer: A

The feed-forward network is applied to each position's vector separately, which adds non-linear capacity after attention mixes information. The attention softmax is the tempting distractor, since it operates on every position, but it mixes information across them. The mask and merge table are not per-position layers.

10. What does the key-value cache store during autoregressive decoding?
- A The full set of training examples used by the model during pretraining runs
 - B The keys and values of earlier tokens in the sequence
 - C The gradients from the most recent backward pass
 - D The output logits of every earlier generation step

Answer: B

Caching previous keys and values lets each new token reuse earlier computations, which speeds generation. Training examples are the tempting distractor, since they are stored in the training pipeline, but the cache is part of inference. Gradients are only produced during training, and earlier output logits are not what the cache reuses.