

PRACTICE QUESTIONS

IABAC

Certified Machine Learning Associate

SHARED BY [u/witty_wombat_57](#) on aiexamcerts

EDITION	Free Set 2
QUESTIONS	10
SHARED	October 15, 2026
EXAM	IABAC Certified Machine Learning Associate

Shared by [u/witty_wombat_57](#) on aiexamcerts

10 practice questions for the IABAC 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/iabac/certified-machine-learning-associate. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. Which encoding is most appropriate for a nominal feature such as country, where no category has a natural order, when the model is linear?
 - A One-hot encoding
 - B Min-max scaling
 - C Target clipping
 - D Ordinal encoding with integers

Answer: A

One-hot encoding creates a binary column per category, so a linear model does not infer a false order between countries. Ordinal integer encoding is the tempting distractor, because it is compact, but it implies one country is numerically greater than another. Min-max scaling applies to numeric values, and target clipping is not an encoding method.

2. R squared computed on held-out data can be negative when a model predicts worse than a constant forecast equal to the mean of the target.
 - A True
 - B False

Answer: A (True)

On unseen data R squared compares the model's squared errors with those of a baseline that always predicts the mean, so a model worse than that baseline scores below zero. Assuming R squared always falls between zero and one is the tempting misreading. The statement is true.

3. A numeric feature with a long right tail, such as transaction amount, needs to be made closer to a bell shape before it is used in a linear model. Which transformation is most commonly applied?
 - A One-hot encoding
 - B Logarithmic transformation
 - C Random oversampling
 - D Principal component analysis

Answer: B

A log transform compresses the long right tail, so very large amounts stop dominating the fit. Principal component analysis is the tempting distractor, since it also transforms numeric inputs, but it builds new uncorrelated axes and does not target skew. One-hot encoding handles categories, and random oversampling addresses class balance.

4. The linear regression variant that adds an L2 penalty on the squared size of coefficients is called _____ regression.

Answer: Ridge

Ridge regression adds the sum of squared coefficients to the loss, which shrinks them and stabilizes estimates when predictors are correlated. Lasso is the tempting answer, since it is the other common penalized linear model, but it uses an L1 penalty. Ordinary least squares applies no penalty at all.

5. Which two statements about the bias and variance trade-off are correct? (Choose two.)

- A High bias usually means the model is too simple to capture the real pattern
- B Adding complexity always lowers both bias and variance
- C High variance usually means the model reacts too strongly to the particular training sample
- D Total error equals bias plus variance with no irreducible noise term
- E Increasing model complexity always reduces variance

Answer: A, C

High bias is underfitting from an overly simple model, and high variance is overfitting to the particular sample. Increasing complexity is the tempting distractor, since complex models often do reduce bias, but they usually raise variance. Total error also includes irreducible noise, and adding complexity does not lower both terms at once.

6. A binary classifier outputs scores. Which curve summarizes its performance across all classification thresholds?

- A The learning curve
- B The elbow curve
- C The residual plot
- D The ROC curve

Answer: D

The ROC curve plots true positive rate against false positive rate at each threshold, and the area under it summarizes ranking quality. The learning curve is the tempting distractor, since it also plots performance, but it tracks error against training size or epochs. The elbow curve helps choose the number of clusters, and a residual plot belongs to regression.

7. Which algorithm classifies a new record by majority vote among the k closest labeled training examples?
- A k-nearest neighbors
 - B Decision tree
 - C Principal component analysis
 - D Gradient boosting

Answer: A

k-nearest neighbors stores the labeled training data and predicts from the most common label among the k nearest examples. Decision tree is the tempting distractor, since it also classifies by splitting data, but it learns rules from splits rather than comparing against stored neighbors. Principal component analysis is unsupervised, and gradient boosting builds an ensemble of trees.

8. Logistic regression produces class probabilities by passing a linear combination of the features through the sigmoid function.
- A True
 - B False

Answer: A (True)

The sigmoid squashes the linear score into a value between zero and one, which is read as the probability of the positive class. Plain linear regression is the tempting distractor, since the model is linear in its inputs, but its outputs are unbounded and are not probabilities. The statement is true.

9. A team uses a validation set to choose a classification threshold. Which practice keeps the final estimate honest?
- A Choosing the threshold on the test set because it is the largest sample
 - B Choosing the threshold from training accuracy
 - C Averaging thresholds across every fold of the test set
 - D Choosing the threshold on the validation set and reporting the untouched test result

Answer: D

Tuning the threshold on validation data and reporting on untouched test data gives an unbiased final figure. Choosing on the test set is the tempting shortcut, since it looks like the most data, but the reported score then is no longer independent. Training accuracy rewards memorization, which makes it a poor basis for any threshold.

10. Which statement about a learning curve is correct?

- A** If training and validation error both stay high as data grows, the model is probably underfitting
- B** It plots accuracy against the number of features in the model
- C** If training error is low and validation error is high, more data is never useful
- D** It always uses the test set

Answer: A

When both curves plateau at poor performance, the model lacks capacity, so more data alone will not fix it. A large gap with low training error points to overfitting, where more data or regularization can help, which makes the claim that they never help wrong. A learning curve plots performance against training set size or epochs, not against features.