

PRACTICE QUESTIONS

IABAC

Certified Machine Learning Associate

SHARED BY [u/wittyrobin2005](#) on aiexamcerts

EDITION	Free Set 1
---------	------------

QUESTIONS	10
-----------	----

SHARED	October 15, 2026
--------	------------------

EXAM	IABAC Certified Machine Learning Associate
------	--

Shared by [u/wittyrobin2005](#) 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. An analyst adds a column recording whether the customer later canceled, then uses it as an input to predict cancellation. What is the main problem?

- A Multicollinearity
- B Target leakage
- C Class imbalance
- D Underfitting

Answer: B

The column holds information that is only known after the outcome, so the model looks accurate in testing but fails in production. This is target leakage. Multicollinearity is the tempting distractor, since it also involves redundant inputs, but it concerns inputs that are correlated with each other rather than with the outcome itself.

2. Extracting the hour of day and the day of week from raw timestamps is an example of _____ engineering.

Answer: feature

Pulling hour and weekday out of a timestamp creates new inputs that expose daily and weekly patterns to the model. Data cleaning is the tempting answer, since timestamps often need cleaning first, but cleaning corrects errors rather than creating new inputs. Hyperparameter tuning adjusts model settings and does not touch the input columns.

3. Which step comes first in a typical machine learning project?

- A Deploying the model to production
- B Tuning hyperparameters
- C Defining the business question and how success will be measured
- D Writing the final evaluation report

Answer: C

A project starts by framing the prediction problem and agreeing on the success metric, because those decide the data needed and what to optimize. Deployment and tuning come much later. Tuning is the tempting distractor, since it is a core modeling step, but it cannot be chosen sensibly until the target and metric are defined.

4. Scaling the full dataset before splitting it into training and test sets is a safe way to avoid leakage, because scaling does not learn anything from the data.
- A True
 - B False

Answer: B (False)

A scaler learns a mean and standard deviation, so fitting it on all rows lets test-set statistics shape the training inputs. Scaling being harmless is the tempting belief. The statement is false, and the scaler should be fitted on the training split only.

5. Which two statements about mean absolute error (MAE) and mean squared error (MSE) are correct? (Choose two.)
- A MAE is expressed in the same units as the target
 - B MAE is always larger than MSE
 - C MSE cannot be used as a training loss
 - D MSE penalizes large errors more heavily than MAE does
 - E Both metrics are unaffected by outliers

Answer: A, D

Squaring errors makes MSE punish large misses more than MAE, and MAE keeps the target's units. Saying MAE is always larger is the tempting distractor, since the two are often compared, but squaring makes MSE larger once errors exceed one unit. MSE is widely used as a training loss, and outliers affect both metrics.

6. The residuals of a regression model plotted against fitted values show a clear curved pattern. What does that suggest?
- A The model is missing a non-linear relationship
 - B The target is predicted perfectly
 - C The learning rate is too small
 - D All features are categorical

Answer: A

Systematic curvature in residuals means the errors depend on the input, so a linear model has missed a non-linear pattern. Random scatter around zero is the healthy picture, and the curve is the tempting thing to ignore. A learning rate or feature type would not produce this particular structure in residuals.

7. Which metric is most appropriate when missing a positive case costs far more than a false alarm, such as failing to detect a serious disease?
- A Precision
 - B Accuracy on the majority class
 - C Recall
 - D Specificity

Answer: C

Recall measures the share of actual positives the model finds, so it is the metric to prioritize when false negatives are the costly error. Precision is the tempting distractor, since it also matters, but it penalizes false alarms rather than missed cases. Accuracy on the majority class can look high while the rare disease goes undetected, and specificity measures correct handling of negatives.

8. The share of predicted positives that are actually positive is called _____.

Answer: precision

Precision divides true positives by the sum of true and false positives, so it focuses on the quality of positive predictions. Recall is the tempting answer, since it also uses true positives, but its denominator is all actual positives. Specificity is about the negative class.

9. What is the main purpose of k-fold cross-validation?
- A To make training faster by using fewer rows
 - B To estimate how well the model generalizes
 - C To select the best feature scaling method only
 - D To replace the need for any test set

Answer: B

Each fold serves once as validation data while the others train the model, and averaging the scores gives a steadier estimate than one split. Faster training is the tempting distractor, since k models are trained, which costs more time, not less. Cross-validation does not remove the need for an untouched final test set when one is required.

10. Which hyperparameter search method evaluates every combination in a specified grid?
- A Random search
 - B Bayesian optimization
 - C Grid search
 - D Stratified sampling

Answer: C

Grid search fits and scores a model for every combination in the parameter grid it is given. Random search is the tempting distractor, since it also tunes hyperparameters, but it samples a fixed number of combinations at random. Bayesian optimization chooses candidates using a surrogate model, and stratified sampling is a data split technique.