

2026-27 CAIO National Qualifier Practical Competition

Part 1 — AI Fundamentals Challenge (Paper A)

Total: 100 marks

Instructions

- Answer all 35 questions.
- Questions 1-15 are single-choice questions (2 marks each). Choose the one best answer.
- Questions 16-25 are multiple-select questions (2 marks each). Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected; otherwise you receive zero for that question.
- Questions 26-35 are fill-in-the-blank questions (5 marks each). Each has exactly five blanks; write your answer for each blank clearly.
- A non-programmable calculator may be used. Total: 100 marks.

Do not open or begin until instructed by the proctor.

Question 1

Data Preparation

2 Points

Consider the following Python code:

```
import numpy as np
A = np.arange(6).reshape(2, 3)
B = A[:, 1:]
B[0, 0] = 100
C = A.flatten()
print(A[0, 1], B[0, 0], C[1], A.sum())
```

What is the printed output?

- A. 100 100 100 114
- B. 1 100 1 114
- C. 100 100 1 15
- D. 1 100 100 114
- E. 100 100 100 15

Question 2

Data Preparation

2 Points

Consider the following Python code:

```
import pandas as pd
df = pd.DataFrame({'k': ['x','y','x','y'], 'v': [10, 20, 30, 40]})
g = df.groupby('k', as_index=False)['v'].mean()
print(g.loc[g['k']=='x', 'v'].values[0], g['v'].sum())
```

What is the printed output?

- A. 20.0 50.0
- B. 20 50
- C. 10.0 50.0
- D. 30.0 50.0
- E. 20.0 100.0

Question 3

Supervised Learning

2 Points

For one-dimensional linear regression the loss is

$$L(w, b) = 1/(2n) \sum_i (y_i - (wx_i + b))^2$$

The gradient of L with respect to w is:

- A. $(2/n)\sum_i x_i(y_i - (wx_i + b))$
- B. $-(1/n)\sum_i x_i(y_i - (wx_i + b))$
- C. $(1/n)\sum_i (y_i - (wx_i + b))$
- D. $-(2/n)\sum_i x_i(y_i - (wx_i + b))$
- E. $(1/n)\sum_i x_i(y_i - (wx_i + b))$

Question 4

Supervised Learning

2 Points

Let X be an $n \times d$ design matrix with $n < d$ (more features than samples) and $\text{rank}(X) = n$.

Which statement is TRUE?

- A. $X^T X$ is invertible, so $\hat{w} = (X^T X)^{-1} X^T y$ is the unique solution.
- B. $X^T X$ is singular; infinitely many w fit the training data exactly, so ridge regularization or the pseudoinverse is needed.
- C. The normal equation has no solution in this case.
- D. The least-squares solution is unique because n equations determine d unknowns.
- E. y must be centered to zero mean before the normal equation can be applied.

Question 5

Supervised Learning

2 Points

With ridge-regularized loss $J(w) = (1/n)\|Xw - y\|^2 + (\lambda/2)\|w\|^2$,

the vectorized gradient-descent update is:

- A. $w := w - \alpha \cdot (2/n)X^T(Xw - y) + \lambda w$
- B. $w := w - \alpha[(2/n)X^T(Xw - y) + \lambda w]$
- C. $w := w - \alpha[(1/n)X^T(Xw - y) + (\lambda/2)w]$
- D. $w := w + \alpha[(2/n)X^T(Xw - y) + \lambda w]$
- E. $w := w - \alpha[(2/n)(Xw - y)X^T + \lambda w]$

Question 6

Supervised Learning

2 Points

For ridge regression $\hat{w}_{\text{ridge}} = (X^T X + \lambda I)^{-1} X^T y$ with $\lambda > 0$,

as $\lambda \rightarrow \infty$:

- A. $\hat{w}_{\text{ridge}}$ approaches the zero vector.
- B. $\hat{w}_{\text{ridge}}$ approaches the ordinary least-squares solution.
- C. the training error approaches zero.
- D. the model becomes equivalent to k-nearest neighbors.
- E. $\|\hat{w}_{\text{ridge}}\|$ grows without bound.

Question 7

Evaluation of ML Models

2 Points

A model has 2% training error and 18% cross-validation error, and the validation curve stays flat as the training size grows.

What is the most appropriate next step?

- A. Add higher-degree polynomial features.
- B. Decrease the regularization parameter λ .
- C. Increase λ (or remove features) to reduce model complexity.
- D. Collect twice as much training data.
- E. Run gradient descent for more iterations.

Question 8

Supervised Learning

2 Points

A decision tree grown to full depth reaches 100% training accuracy but only 70% test accuracy.

Which change is LEAST likely to improve test accuracy?

- A. Pruning the tree with cost-complexity pruning.
- B. Setting `max_depth` to a smaller value.
- C. Requiring `min_samples_leaf` ≥ 10 .
- D. Growing the tree even deeper.
- E. Collecting more training data.

Question 9

Unsupervised Learning

2 Points

Which statement about the k-means algorithm is TRUE?

- A. It always finds the globally optimal clustering.
- B. The number of clusters k must be chosen in advance, and different random initializations can converge to different local optima.
- C. It works well when clusters are elongated or non-spherical.
- D. Its result is insensitive to feature scaling.
- E. It automatically determines the best number of clusters.

Question 10

Unsupervised Learning

2 Points

You fit a Gaussian to each feature for anomaly detection and pick the threshold ϵ by maximizing F1 on a cross-validation set (flag x as anomaly if $p(x) < \epsilon$).

If ϵ is set far too large, what happens?

- A. Only the most extreme anomalies are flagged; recall collapses.

- B. Many normal points are flagged as anomalies, so precision drops.
- C. The fitted Gaussian parameters μ and σ^2 change.
- D. Recall decreases while precision increases.
- E. Nothing changes; ϵ only affects training time.

Question 11

Evaluation of ML Models

2 Points

A binary classifier's confusion matrix: TP = 80, FP = 20, FN = 30, TN = 870.

What are precision, recall, and F1 (rounded to two decimals)?

- A. Precision 0.80, Recall 0.73, F1 $\approx$ 0.76
- B. Precision 0.73, Recall 0.80, F1 $\approx$ 0.76
- C. Precision 0.80, Recall 0.73, F1 $\approx$ 0.66
- D. Precision 0.89, Recall 0.80, F1 $\approx$ 0.84
- E. Precision 0.80, Recall 0.97, F1 $\approx$ 0.88

Question 12

Kernel Methods

2 Points

Consider a trained hard-margin support vector machine.

Removing which training points can change the learned decision boundary?

- A. Only the support vectors.
- B. All training points influence it equally.
- C. Only the misclassified points.
- D. No single training point can change it.
- E. Only the points farthest from the boundary.

Question 13

Kernel Methods

2 Points

For the RBF kernel $k(x, z) = \exp(-\gamma \|x - z\|^2)$,

increasing γ typically:

- A. makes the boundary smoother and reduces overfitting.
- B. makes the boundary more complex and localized; the risk of overfitting grows.
- C. turns the kernel into a linear kernel.
- D. makes training faster.
- E. always reduces the number of support vectors.

Question 14

AI Search

2 Points

True cheapest costs to the goal: $h^*(S) = 8$, $h^*(A) = 5$, $h^*(B) = 6$.

Which heuristic is admissible (never overestimates)?

- A. $h(S)=9$, $h(A)=4$, $h(B)=6$
- B. $h(S)=8$, $h(A)=5$, $h(B)=7$
- C. $h(S)=7$, $h(A)=5$, $h(B)=5$
- D. $h(S)=8$, $h(A)=6$, $h(B)=6$
- E. $h(S)=10$, $h(A)=5$, $h(B)=6$

Question 15

Logical Reasoning

2 Points

Knowledge base $KB = \{ P \rightarrow Q, \neg Q \}$.

Which formula is logically entailed by KB?

- A. P
- B. Q
- C. $\neg P$
- D. $P \wedge Q$
- E. $P \vee Q$

Question 16

Data Preparation

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which of the following snippets execute WITHOUT raising an error?

- A. `np.array([1, 2, 3]) + np.array([[1], [2]])`
- B. `pd.DataFrame({'a': [1, 2]}).['b']`
- C. `np.zeros((2, 3)).reshape(3, 2)`
- D. `np.arange(5)[[0, 2, 4]]`
- E. `pd.DataFrame({'a': [1, 2, 3]}).['a'][5]`

Question 17

Supervised Learning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about gradient descent and diagnostics are TRUE?

- A. Batch gradient descent computes the gradient over the entire training set at each step.

- B. The noisy updates of SGD can help it escape shallow local minima more easily than batch GD.
- C. Vectorizing across examples mainly reduces the number of iterations needed to converge.
- D. A large, persistent gap between training and cross-validation error indicates high variance.
- E. Decreasing the learning rate α always speeds up convergence.

Question 18

Supervised Learning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about L1 vs L2 regularization are TRUE?

- A. L1 regularization tends to produce sparse weight vectors (many exact zeros).
- B. L2 regularization shrinks weights toward zero but rarely sets them exactly to zero.
- C. The L1 penalty is differentiable at $w = 0$.
- D. Ridge regression has a closed-form solution, while lasso generally does not.
- E. Increasing λ always reduces the training error.

Question 19

Supervised Learning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about tree-based models are TRUE?

- A. Random forests reduce variance by averaging trees trained on bootstrap samples with random feature subsets.
- B. Gradient boosting primarily reduces bias by sequentially fitting the residual errors.
- C. Increasing `max_depth` of a single decision tree tends to increase variance.
- D. In XGBoost, a larger learning rate η always improves generalization.
- E. Pruning a decision tree increases its complexity.

Question 20

Unsupervised Learning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about clustering are TRUE?

- A. k-means implicitly assumes roughly spherical clusters and is sensitive to feature scaling.
- B. DBSCAN can discover arbitrarily shaped clusters and label sparse points as noise.
- C. Agglomerative hierarchical clustering requires the number of clusters k up front,

like k-means.

D. A Gaussian mixture model assigns every point to exactly one cluster.

E. k-means++ initialization reduces the risk of poor local optima compared with uniform random init.

Question 21

Evaluation of ML Models

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about evaluation metrics are TRUE?

- A. On a dataset with 99% negatives, always predicting 'negative' scores 99% accuracy.
- B. Precision-recall curves are more informative than ROC curves for highly imbalanced data.
- C. ROC-AUC is independent of the chosen classification threshold.
- D. A higher R^2 on the training set always means better predictions on new data.
- E. Micro-averaged F1 gives equal weight to each class regardless of its size.

Question 22

Kernel Methods

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about SVMs and kernels are TRUE?

- A. The kernel trick computes inner products in the feature space without explicitly mapping the data.
- B. The RBF kernel corresponds to an infinite-dimensional feature space.
- C. Only the support vectors determine the SVM decision boundary.
- D. SVMs are insensitive to feature scaling.
- E. A linear kernel can separate every dataset that an RBF kernel can separate.

Question 23

AI Search

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about search algorithms are TRUE?

- A. BFS is complete and optimal when all step costs are equal.
- B. DFS uses $O(bm)$ memory but is neither optimal nor complete in infinite spaces.
- C. Uniform-cost search is optimal when all step costs are positive.
- D. A* guarantees an optimal solution regardless of the heuristic used.
- E. Greedy best-first search is optimal.

Question 24

Recommender Systems

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about recommender systems are TRUE?

- A. User-based collaborative filtering recommends items liked by users with similar tastes.
- B. Item-based filtering is usually more scalable when users far outnumber items.
- C. Collaborative filtering has no cold-start problem.
- D. TF-IDF features of item descriptions can be used in content-based recommendation.
- E. Pearson correlation always yields the same neighbor ranking as cosine similarity.

Question 25

Reinforcement Learning

2 Points

Select ALL correct options. You receive full marks only if every correct option — and no incorrect option — is selected.

Which statements about reinforcement learning are TRUE?

- A. In ϵ -greedy action selection, the agent exploits the best-known action with probability $1-\epsilon$.
- B. Q-learning is an off-policy temporal-difference method updating toward $\max_a Q(s', a)$.
- C. A discount factor γ close to 0 makes the agent far-sighted.
- D. The Bellman optimality equation expresses $V^*(s)$ as the maximum expected return over actions.
- E. Increasing the exploration rate ϵ always improves the agent's final return.

Question 26

Data Preparation

5 Points

Data Cleaning and Aggregation (fill in the 5 blanks)

orders.csv columns: order_id, customer, date (string 'YYYY-MM-DD'),

amount (float, may be NaN), region. Complete the script:

```
import pandas as pd
df = pd.read_csv('orders.csv')
df['date'] = pd._____(df['date'])
df = df.dropna(_____)
df['month'] = df['date'].dt.month
g = df._____(['region', 'month'])['amount']._____( )
g = g._____(ascending=False)
```

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____
- (d) _____
- (e) _____

Question 27

Supervised Learning

5 Points

Logistic Regression Without Data Leakage (fill in the 5 blanks)

students.csv columns: hours, score, passed (1/0). Complete the script. Watch out: the scaler must not see the test set.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
df = pd.read_csv('students.csv')
X = df[['hours', 'score']]; y = df['passed']
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=_____)
scaler = StandardScaler()
X_train_s = scaler._____(X_train)
X_test_s = scaler._____(X_test)
model = LogisticRegression()
model._____(X_train_s, y_train)
print(_____(y_test, model.predict(X_test_s)))
```

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____
- (d) _____
- (e) _____

Question 28

Supervised Learning

5 Points

Information Gain, Step by Step (fill in the 5 blanks; 2 decimals)

12 samples: Positive 7, Negative 5. Split on X:

Left: P 5, N 1. Right: P 2, N 4. (log base 2)

Hints: $\log_2(7/12) = -0.778$, $\log_2(5/12) = -1.263$, $\log_2(5/6) = -0.263$,

$\log_2(1/6) = -2.585$, $\log_2(2/6) = -1.585$, $\log_2(4/6) = -0.585$.

H(parent) $\approx$ _____
 H(left) $\approx$ _____
 H(right) $\approx$ _____
 weighted conditional entropy $\approx$ _____
 information gain $\approx$ _____

Write your answer for each blank below:

- (a) _____
 (b) _____
 (c) _____
 (d) _____
 (e) _____

Question 29

Supervised Learning

5 Points

Vectorized Gradient and Update (fill in the 5 blanks)

Complete the vectorized gradient of $J(w,b) = (1/2m)\|Xw+b\cdot 1-y\|^2$

(X: $m \times n$, y: m-vector) and one gradient-descent step with learning rate alpha.

```
import numpy as np
def compute_gradient(X, y, w, b):
    m, n = X.shape
    f_wb = _____
    err = _____
    dj_dw = _____
    dj_db = _____
    return dj_dw, dj_db
# one GD step:
w = _____
```

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____
- (d) _____
- (e) _____

Question 30

Supervised Learning

5 Points

Logistic Regression by Hand (fill in the 5 blanks)

Trained model: $w = [2, -1]$, $b = -0.5$. Hint: $e^{0.5} \approx 1.649$, $e^{3.5} \approx 33.1$.

Round probabilities to two decimals.

For $x = (1, 2)$: $z = w \cdot x + b =$ _____
 $\sigma(z) \approx$ _____
 predicted class (threshold 0.5) = _____
 For $x = (0, 3)$: $z =$ _____ ; $\sigma(z) \approx$ _____

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____
- (d) _____
- (e) _____

Question 31

Unsupervised Learning

5 Points

One k-means Iteration (fill in the 5 blanks)

Points: (1,1), (1,2), (4,4), (5,4). Initial centroids $\mu_1=(1,1)$, $\mu_2=(5,5)$.

Do ONE Lloyd iteration (assign, then recompute).

new $\mu_1 = (\text{_____, _____})$ # x, y
 new $\mu_2 = (\text{_____, _____})$ # x, y
 After this iteration the algorithm has converged (yes/no): _____

Write your answer for each blank below:

- (a) _____
- (b) _____
- (c) _____

(d) _____

(e) _____

Question 32

AI Search

5 Points

A* Trace (fill in the 5 blanks)

Edges: $S \rightarrow A$ cost 1, $S \rightarrow B$ cost 2, $A \rightarrow G$ cost 6, $B \rightarrow G$ cost 3.Heuristic: $h(S)=5$, $h(A)=5$, $h(B)=2$, $h(G)=0$.After expanding S: $f(A) = \underline{\hspace{2cm}}$, $f(B) = \underline{\hspace{2cm}}$

The node expanded next (after S) is _____

 $f(G)$ via the next expansion = _____The optimal path found is _____ (e.g. $S \rightarrow B \rightarrow G$)

Write your answer for each blank below:

(a) _____

(b) _____

(c) _____

(d) _____

(e) _____

Question 33

Kernel Methods

5 Points

RBF Kernel by Hand (fill in the 5 blanks; 3 decimals)

 $k(x,z) = \exp(-\gamma \|x-z\|^2)$. Hint: $e^{-2.5} \approx 0.082$, $e^{-0.9} \approx 0.407$. $x=(1,2)$, $z=(4,6)$, $\gamma=0.1$: $\|x-z\|^2 = \underline{\hspace{2cm}}$ $-\gamma \|x-z\|^2 = \underline{\hspace{2cm}}$; $k(x,z) \approx \underline{\hspace{2cm}}$ $x=(1,2)$, $z=(1,5)$, $\gamma=0.1$: $\|x-z\|^2 = \underline{\hspace{2cm}}$; $k(x,z) \approx \underline{\hspace{2cm}}$

Write your answer for each blank below:

(a) _____

(b) _____

(c) _____

(d) _____

(e) _____

