AI Olympiad, September 2024

Please note that if you have questions you are free to ask them in your mother tongue through the messaging system.

Marks for each part of each question are indicated in square brackets.

Question 1

1. Assume a model for medical diagnosis uses "accuracy" for evaluation. Given that only 5% (ground truth) of patients have a rare disease, this evaluation metric is considered problematic.

   (a) Give a general explanation of why and

   [1 Marks]

   (b) a specific case for the medical diagnosis model that demonstrates this.

   [1 Marks]

2. Consider a spam email detection system where only 1% (ground truth) of emails are spam.

   (a) Which evaluation metric is considered the best choice: "precision," "recall," or "f1-score"?

   [1 Marks]

   (b) Explain the benefits of the best metric.

   [2 Marks]

   (c) Criticize the remaining metrics.

   [2 Marks]

[Total 7 Marks]

Question 2

Note that in this question marks for each part will be awarded only if all of the good answers have been identified.

1. A trolley is heading towards a mystery bus with a 50% chance of containing two people. You can pull the lever to divert it to the other track, hitting a mystery bus with a 10% chance of 10 people instead. What should you do (select all good answers)?

   A. Pull the lever because it saves more lives.

   B. Don't pull the lever because it saves more lives.

   C. Neither option saves more lives.

   D. It is a moral dilemma. It shouldn't be solved with a probability equation.

   [1 Marks]

2. The Biased Job Applicant: An AI system used by a company to screen job applicants favors candidates from a particular socioeconomic background due to historical data. To address fairness, the company must (select all good answers).

   A. Ignore the bias since the AI reflects past decisions, which may be considered fair.

   B. Adjust the AI to remove the bias, but risk losing historical consistency.

   C. Collect new, more diverse data and retrain the AI, potentially delaying hiring processes.

   D. Use a combination of adjustments and retraining, while considering the cost and time implications.

   [1 Marks]

3. The Ethical Self-Driving Car: In an unavoidable accident, an autonomous vehicle must choose between hitting an elderly pedestrian or a young child. The AI should decide by (select all good answers).

   A. Choose to hit the elderly pedestrian, considering the potential remaining life span.

B. Choose to hit the young child, considering the likelihood of faster recovery.

C. Follow programmed ethical guidelines that prioritize minimizing overall harm.

D. Make a random decision to avoid bias.

[1 Marks]

4. A hospital's AI is less accurate in diagnosing women due to male-biased training data. Addressing this requires (select all good answers):

   A. Adjusting the AI's algorithms to improve accuracy for women, risking unforeseen side effects.

   B. Retraining the AI with a more balanced dataset, which may delay its use.

   C. Using a combination of retraining and adjustments, which might increase costs and complexity.

   D. Continuing to use the AI since it works well for men, potentially compromising women's health.

[1 Marks]

5. The Surveillance Dilemma: AI-powered surveillance cameras disproportionately target certain racial or ethnic groups. To address fairness, the city should (select all good answers).

   A. Modify the AI to eliminate racial or ethnic bias, which may affect its effectiveness.

   B. Implement oversight to ensure fair use of the surveillance system, requiring additional resources.

   C. Keep the system as is since it reflects crime data, potentially perpetuating biases.

   D. Use a combination of modifications and oversight, balancing fairness and resource constraints.

[1 Marks]

6. The Transparent Algorithm: A social media platform's AI algorithm is suspected of promoting specific political views. To address this, the company could (select all good answers):

   A. Not disclose how the algorithm works to protect proprietary information, risking public distrust.

   B. Disclose fully to ensure transparency and build user trust, potentially exposing proprietary methods.

   C. Partially disclose to balance transparency and proprietary interests, possibly causing confusion.

   D. Conduct independent audits instead of full disclosure, maintaining some level of secrecy.

[1 Marks]

7. What is a possible ethical resolution to address fair use of copyrighted material in AI training (select all good answers)?

   A. Only using data that is explicitly labeled as free to use.

   B. Negotiating agreements with content owners for data usage.

   C. Implementing a system to track and credit original content creators.

   D. Limiting AI training to non-commercial data sources.

[1 Marks]

[Total 7 Marks]

Question 3

Consider the following 2D dataset:

Data Point #	x	y
1	1.90	0.97
2	1.76	0.84
3	2.32	1.63
4	2.31	2.09
5	1.14	2.11
6	5.02	3.02
7	5.74	3.84
8	2.25	3.47
9	4.71	3.60
10	3.17	4.96

Suppose the initial assignment of cluster centers based on (x, y) coordinates are:

$$\theta_{A,0} : (1.90, 0.97), \quad \theta_{B,0} : (3.17, 4.96)$$

Assuming k-means uses Euclidean distance,

$$d(p, q) = \|p - q\|_2^2 = \sqrt{\sum_{i=1}^{d}(p_i - q_i)^2}$$

1. Simulate the k-means (k=2) algorithm cluster assignment. What are the cluster assignments and distances from the nearer of the initial centers $\theta_{A,0}$ and $\theta_{B,0}$ after cluster assignment?

[2 Marks]

Data #	Cluster Assignment	Distance from the Cluster Centre
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

2. Based on the previous cluster assignment, we now calculate the new cluster centers $\theta_{A,1}$ and $\theta_{B,1}$. What are the new coordinates of $\theta_{A,1}$ and $\theta_{B,1}$? Make the steps in your computation explicit.

[2 Marks]

3. Let a configuration of the k-means algorithm correspond to the k-way partition (on the set of instances to be clustered) generated by the clustering at the end of each iteration. Is it possible for the k-means algorithm to revisit a configuration? Justify your answer and show why this proves that the k-means algorithm converges in a finite number of steps.

[3 Marks]

[Total 7 Marks]

Question 4

DALL-E uses a discrete VAE (Variational Autoencoder) to encode images into tokens and then generates images from these tokens using a Transformer architecture. Suppose the model uses a vocabulary of $V = 8192$ discrete tokens. Each token is represented by an embedding vector of dimensionality $d = 512$.

1. Calculate the total number of parameters in the embedding matrix used for encoding these tokens.

   [3 Marks]

2. The feed-forward network within each Transformer layer consists of two linear transformations: one with input dimension d and output dimension $f = 2048$, and another with input dimension f and output dimension d. Compute the total number of parameters for the feed-forward network in a single Transformer layer.

   [4 Marks]

[Total 7 Marks]

Question 5

1. The Perceptron algorithm applied to the training set

   $$((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m))$$

   involves the following training loop:

   ```
   while not converged do
     select $i$ from $\{1, \ldots, m\}$
     if $y_i\langle \mathbf{w}, \mathbf{x}_i\rangle \leq 0$ then
       $\mathbf{w} \leftarrow \mathbf{w} + y_i\mathbf{x}_i$;
     end
   end
   ```

   if $\mathbf{w}$ is initialised to 0, show that $\mathbf{w}$ can be expressed as a linear combination:

   $$\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i,$$

   of the training data, explaining what the value of α_i will be at any stage of the algorithm.

   [1 Marks]

2. Show we can evaluate the classification $\operatorname{sgn}(\langle \mathbf{w}, \mathbf{x}\rangle)$ on a test example $\mathbf{x}$ using the values of α_i and inner products between $\mathbf{x}$ and the training data $\langle \mathbf{x}, \mathbf{x}_i\rangle$, for $i = 1, \ldots, m$. Hence, show how we can run the Perceptron algorithm in a feature space defined by a mapping

   $$\phi : \mathbf{x} \longmapsto \phi(\mathbf{x})$$

   using only the kernel function

   $$\kappa(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z})\rangle.$$

   [2 Marks]

3. Novikoff's theorem guarantees convergence provided there is a weight vector that correctly classifies the training data. Assuming the training examples are distinct, consider the modification of a normalised kernel κ to

   $$\tilde{\kappa}(\mathbf{x}, \mathbf{z}) = \kappa(\mathbf{x}, \mathbf{z}) + a\delta(\mathbf{x}, \mathbf{z}),$$

   where $a > 0$ and

   $$\delta(\mathbf{x}, \mathbf{z}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{z}, \\ 0 & \text{otherwise.} \end{cases}$$

Show we can find a vector α such that $\tilde{K}\alpha = \mathbf{y}$ for the label vector

$$\mathbf{y} = (y_1, \ldots, y_m)^T,$$

where $\tilde{K}$ is the kernel matrix of the modified kernel $\tilde{\kappa}$. Hence, argue that the training set is separable in the feature space defined by the mapping $\tilde{\phi}$ of the kernel $\tilde{\kappa}$.

[2 Marks]

4. Give an upper bound on the number of updates the Perceptron algorithm will make using the kernel $\tilde{\kappa}$ in terms of the vector α from part 3.

[2 Marks]

[Total 7 Marks]

Question 6

This question focuses on evaluating Markov Decision Processes (MDPs) and Bellman Equations in reinforcement learning. Assume an agent is navigating a grid world with the following characteristics:

- The grid is a 3x3 matrix with each cell representing a state $s \in \{s_1, s_2, \ldots, s_9\}$:

s_1	s_2	s_3
s_4	s_5	s_6
s_7	s_8	s_9

- The agent can move up, down, left, or right. If the agent tries to move outside the grid, it stays in the same position.

- The rewards are given as follows:
  - $R(s_1, \text{attempts right}) = 1$
  - $R(s_2, \text{attempts down}) = 2$
  - All other rewards are 0.

- The transition probabilities are stochastic:
  - With probability 0.7, the agent actually moves in the chosen direction.
  - With probability 0.1, the agent actually moves left.
  - With probability 0.1, the agent actually moves right.
  - With probability 0.1, the agent actually moves down.

The agent starts at s_1. Assume a discount factor of $\gamma = 0.9$.

1. Consider performing Q-learning with the value function V initialised to 0. Calculate the expected state-action value $Q(s_1, \text{right})$ after attempting to move right in the first iteration. [1 Mark]

2. Derive the Bellman equation for state s_1 assuming the agent uses a policy π that always chooses the action "attempt right" in s_1 and "attempt down" in all other states. [1 Mark]

3. Compute the value function V_π for the policy π. [5 marks]

[Total 7 Marks]

End Of Paper

Model Answers

Question 1

1. [a)]

   Accuracy can be misleading on imbalanced datasets where one class is significantly more frequent than the other. A high accuracy may not reflect the model's performance adequately as it could simply be predicting the majority class most of the time.

   1 mark for identifying issue.

(b) In the medical diagnosis scenario where only 5% of patients have a rare disease, if the model always predicts "no disease," it could achieve 95% accuracy but would be useless for identifying the disease.

   1 mark for giving clear demonstration of application.

2. [a)]

   F1-score

(b) F1-score balances precision and recall, making it suitable for imbalanced datasets. It considers both false positives (FP) and false negatives (FN). In the spam email detection system, f1-score therefore provides a single metric to evaluate how well the model identifies spam while minimizing false alarms.

   1 mark for pointing out F1 balancing prec and recall or FP and FN, plus 1 mark for implications for spam detection.

(c) Precision and recall focus on specific classes and can also be biased by class imbalance. High precision may be achieved by simply predicting the majority class while recall may suffer for the minority class. In the spam email detection system where spam emails are rare (1%), a model might have high precision (few false positives) but very low recall (misses many actual spam cases).

   similarly 1 mark for limitations of precision and recall and 1 for the implications for spam.

Question 2

1. In this scenario, you need to consider the expected value of the potential outcomes to make a decision.

    - **Track A (no action):**
        - 50% chance of hitting 2 people: $0.5 \times 2 = 1$ person
        - 50% chance of hitting 0 people: $0.5 \times 0 = 0$ people
        - Expected value: $1 + 0 = 1$ person
    - **Track B (pull the lever):**
        - 10% chance of hitting 10 people: $0.1 \times 10 = 1$ person
        - 90% chance of hitting 0 people: $0.9 \times 0 = 0$ people
        - Expected value: $1 + 0 = 1$ person

    The expected value for both tracks is the same, which is 1 person. Given that the expected value is equal, other ethical considerations might come into play. Correct answer C,D

2. Correct answer B,C,D

3. Correct answer A,C,D

4. Correct answer A,B,C

5. Correct answer A,B,D

6. Correct answer B,C,D

7. Correct answer A,B,C

Question 3

1.

Data #	Cluster Assignment	Distance from the Cluster Centre
1	1	0
2	1	$\sqrt{(1.76-1.90)^2+(0.84-0.97)^2} = 0.191$
3	1	0.7823
4	1	1.1929
5	1	1.37011
6	2	2.68069
7	2	2.80344
8	2	1.75114
9	2	2.05456
10	2	0

1 mark for the cluster assignment and 1 for the distances - potential of one or two errors in distances being tolerated.

2.

$$\theta_{A,1} = \left(\frac{x_1+x_2+x_3+x_4+x_5}{5}, \frac{y_1+y_2+y_3+y_4+y_5}{5}\right) = (1.886, 1.528)$$

$$\theta_{B,1} = \left(\frac{x_6+x_7+x_8+x_9+x_{10}}{5}, \frac{y_6+y_7+y_8+y_9+y_{10}}{5}\right) = (4.178, 3.778)$$

One point for the formula and a second for the completed computation.

3. Since the k-means algorithm converges if the k-way partition does not change in successive iterations, thus the k-way partition has to change after every iteration. As the mean squared error monotonically decreases, it is thus impossible to revisit a configuration. Therefore, the k-means algorithm will eventually run out of configurations and converge.

1 mark for argument that partition decides state of algorithm, so finite number. Any change reduces average squared error so must converge.

Question 4

1. Total Parameters $= V \times d = 8192 \times 512 = 4,194,304$

   2 marks for right formula, 1 for correct computation.

2. The feed-forward network consists of two linear transformations:

   - From d to f.
   - From f back to d (input dimension f and output dimension d).

   Parameters for First Linear Transformation $= d \times f = 512 \times 2048 = 1,048,576$

   Parameters for Second Linear Transformation $= f \times d = 2048 \times 512 = 1,048,576$

   Total Parameters for Feed-Forward Network $= 1,048,576 + 1,048,576 = 2,097,152$

   2 marks for understanding the two transformations. 2 for computing correctly.

Question 5

1. Initially $\alpha_i = 0$ for all i. If $\mathbf{w} \leftarrow \mathbf{w} + y_i \mathbf{x}_i$ we get the new $\mathbf{w}$ by incrementing α_i by 1. Hence, α_i counts the number of times $(\mathbf{x}_i, y_i)$ has caused the weight to be updated.

   1 mark for identifying that can update α_i for each iteration.

2. $$\begin{aligned} sgn(\langle \mathbf{w}, \mathbf{x} \rangle) &= \operatorname{sgn}\left(\left\langle \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i, \mathbf{x} \right\rangle\right) \\ &= \operatorname{sgn}\left(\sum_{i=1}^m \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x} \rangle\right) \end{aligned}$$

   Hence,

   $$\operatorname{sgn}(\langle \mathbf{w}, \phi(\mathbf{x}) \rangle) = \operatorname{sgn}\left(\sum_{i=1}^m \alpha_i y_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle\right) = \operatorname{sgn}\left(\sum_{i=1}^m \alpha_i y_i \kappa\left(\mathbf{x}_i, \mathbf{x}\right)\right)$$

   This also means we can evaluate the test expression in the basic training loop, while the update for α is given in the answer to (a).

   1 mark for observing the right expression, 1 mark for correct derivation.

3. $$\tilde{K} = K + aI$$

   is non singular as $a > 0$ and K is positive semi-definite, and so there exists α such that $\tilde{K}\alpha = \mathbf{y}$. Setting

   $$\mathbf{w} = \sum_{i=1}^m \alpha_i \tilde{\phi}(\mathbf{x}_i)$$

   we have a separating hyperplane.

   1 mark for observing that the matrix is non-singular. 1 mark for inferring that solution exists.

4. The margin of the hyperplane determined by α is $\gamma = \|\mathbf{w}\|^{-1} = 1/\sqrt{\alpha^T \tilde{K} \alpha}$, since we need to normalise the weight vector. Since the radius R of the ball containing the data is $\sqrt{\tilde{\kappa}(\mathbf{x}, \mathbf{x})} = \sqrt{1+a}$, the bound given by the Perceptron convergence algorithm is

   $$\frac{R^2}{\gamma^2} = (1+a)\alpha^T \tilde{K} \alpha = (1+a)\mathbf{y}^T \alpha = (1+a)\mathbf{y}^T \tilde{K}^{-1} \mathbf{y} \leq \frac{m(1+a)}{a}.$$

   Note: full marks if first equality given in last equation.

   1 mark for getting approximately right (eg wrong norm or estimate of γ).

Question 6

1. $$Q(s_1, \text{right}) = R(s_1, \text{right}) + \gamma \sum_{s'} P(s'|s_1, \text{right}) V(s')$$

   For s_1:

   $P(s_2|s_1, \text{right}) = 0.7{+}0.1, \quad P(s_1|s_1, \text{right}) = 0.1 \quad P(s_4|s_1, \text{right}) = 0.1$

   $$Q(s_1, \text{right}) = 1 + 0.9(0.8 \times V(s_2) + 0.1 \times V(s_1) + 0.1 \times V(s_4))$$

   Since initial values of V are 0:

   $$Q(s_1, \text{right}) = 1$$

   1 mark if correct expression - also mark if correct formula but error in evaluation.

2. The Bellman equation for s_1 under policy π:

   $$V(s_1) = \sum_a \pi(a|s_1) \left(R(s_1, a) + \gamma \sum_{s'} P(s'|s_1, a) V(s') \right)$$

   Since π always chooses "right" in s_1:

   $$V(s_1) = R(s_1, \text{right}) + \gamma(0.8V(s_2) + 0.1V(s_1) + 0.1V(s_4))$$

   1 mark for correct expression.

3. Solving the Bellman equations with π: first observe that $V(s_i) = 0$ for all $i \geq 0$, since it is impossible to regain the top row and rewards are only accessible from the top row. Hence, need only solve for s_1, s_2 and s_3. Have simultaneous equations:

   $$\begin{aligned} V(s_1) &= 1 + 0.9(0.1 \times V(s_1) + 0.8 \times V(s_2)) \\ V(s_2) &= 2 + 0.9(0.1 \times V(s_1) + 0.1 \times V(s_3)) \\ V(s_3) &= 0.9(0.1 \times V(s_3) + 0.1 \times V(s_2)) \end{aligned}$$

   3 marks for observing this far. An extra mark for writing as a matrix equation:

   $$\begin{pmatrix} 0.91 & -0.72 & 0 \\ -0.09 & 1 & -0.09 \\ 0 & -0.09 & 0.91 \end{pmatrix} \begin{pmatrix} V(s_1) \\ V(s_2) \\ V(s_3) \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \tag{1}$$

and final mark for providing a solution:

$$\begin{pmatrix} 0.91 & -0.72 \\ -0.0819 & 0.9019 \end{pmatrix} \begin{pmatrix} V(s_1) \\ V(s_2) \end{pmatrix} = \begin{pmatrix} 1 \\ 1.82 \end{pmatrix} \tag{2}$$

$$V(s_1) = \frac{0.9019 + 1.82 \times 0.72}{0.9019 \times 0.91 - 0.72 \times 0.0819} = 2.904191735$$

$$V(s_2) = \frac{0.0819 + 0.91 \times 1.82}{0.9019 \times 0.91 - 0.72 \times 0.0819} = 2.281686776$$

$$V(s_3) = \frac{V(s_2) \times 0.09}{0.91} = 0.225661329$$

If result obtained by calculation then can obtain marks if result is correct, otherwise partial marks if algorithmic approach partially correct.

End Of Paper