

CAP 5636 Advanced Artificial Intelligence

Sample Exam Problems with Solutions

Fall 2026

These problems are taken from CAP 5636 exams of previous years. They show the kinds of questions asked and the expected form of the answers. The instructions, scope, and format of the 2026 exams are announced separately.

Part I contains the problems; Part II contains their solutions and point breakdowns. Try the problems before reading the solutions.

Part I: Problems

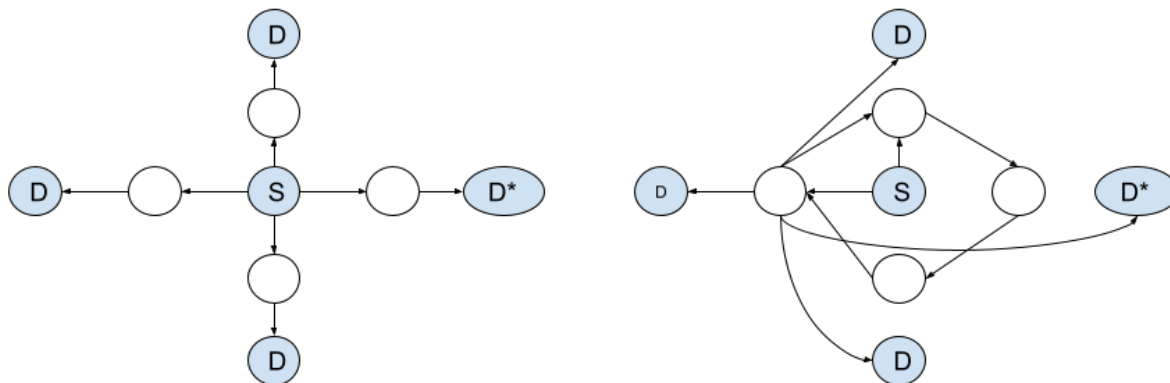
Problem 1: Joint probability queries

Consider the following joint probability table:

Sneezing	Fever	Sinus-pain	Allergy	P
not sneezing	no fever	no pain	no allergy	0.252
not sneezing	no fever	no pain	allergy	0.252
not sneezing	no fever	pain	no allergy	0.028
not sneezing	no fever	pain	allergy	0.028
not sneezing	fever	no pain	no allergy	0.063
not sneezing	fever	no pain	allergy	0.063
not sneezing	fever	pain	no allergy	0.007
not sneezing	fever	pain	allergy	0.007
sneezing	no fever	no pain	no allergy	0.0216
sneezing	no fever	no pain	allergy	0.1944
sneezing	no fever	pain	no allergy	0.0024
sneezing	no fever	pain	allergy	0.0216
sneezing	fever	no pain	no allergy	0.0054
sneezing	fever	no pain	allergy	0.0486
sneezing	fever	pain	no allergy	0.0006
sneezing	fever	pain	allergy	0.0054

1. Calculate the probability that someone has allergy, $P(\text{allergy})$. Explain the process of calculation. (2 points)
2. Calculate the probability that someone has allergy given they are sneezing, $P(\text{allergy} \mid \text{sneezing})$. Explain the process of calculation. (1 point)
3. Calculate the probability that someone has allergy given they have fever but are not sneezing, $P(\text{allergy} \mid \text{fever, not sneezing})$. Explain the process of calculation. (1 point)
4. Are the probabilities of pain and fever independent of each other (i.e., is $\text{Sinus-pain} \perp\!\!\!\perp \text{Fever}$)? Explain how you reached your answer. (1 point)

Problem 2: Search algorithm comparison



Consider the two search problems Left and Right above. Assume that the locations sit on a 1×1 grid. Every edge represents an action, is directional, and has a cost of 1. The location labeled **S** is the start location; the locations labeled **D** and **D*** are goals for uninformed search. The location labeled **D*** is the goal for informed search.

For all the algorithms, assume that the successors of a node are expanded clockwise, starting from 12 o'clock, according to the direction from the node to the successor's location on the grid. For example, the node at 12 o'clock comes first, then a node at 1 o'clock, then one at 1:30, then the nodes at 3 o'clock, 6 o'clock, and 9 o'clock. For a curved edge, use the direction to the successor's location, not the shape of the arrow.

Discuss how well each of the following algorithms performs for the Left and Right problems. Specify whether it finds a solution, whether it finds the optimal solution, and how efficient it is in the search. For graph search, discuss whether graph search saves some expansions. For A*, state whether it is *guaranteed* to find the optimal solution, and justify your answer by discussing whether the heuristic is admissible and consistent.

- 1) Breadth-first tree search
- 2) Breadth-first graph search
- 3) Depth-first tree search
- 4) Greedy search with the Euclidean distance as heuristic
- 5) A* search with the Euclidean distance as heuristic

List your answers as follows:

1-Left
 2-Left

 1-Right

Problem 3: Connect Four evaluation function

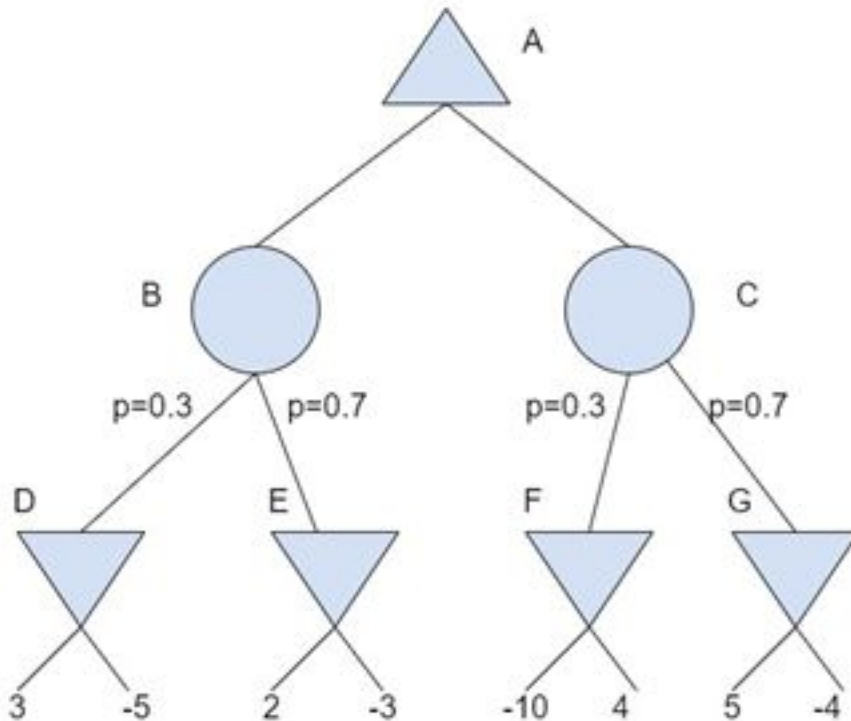


Connect Four is a two-player board game in which the players take turns dropping colored tokens into a seven-column, six-row vertically suspended grid. The pieces fall straight down, occupying the lowest available space within the column. The objective of the game is to be the first to form a horizontal, vertical, or diagonal line of four of one's own tokens.

- What is the branching factor b of the game? Discuss.
- Propose an evaluation function for the game.
- Estimate the computational complexity of the proposed function (assume m rows and n columns). For the game as sold, $m = 6$ and $n = 7$, but you can imagine different sizes.
- Explain in one sentence why this is an appropriate evaluation function.
- Explain in one sentence why this is not a perfect evaluation function (or why it is perfect, if it is perfect).

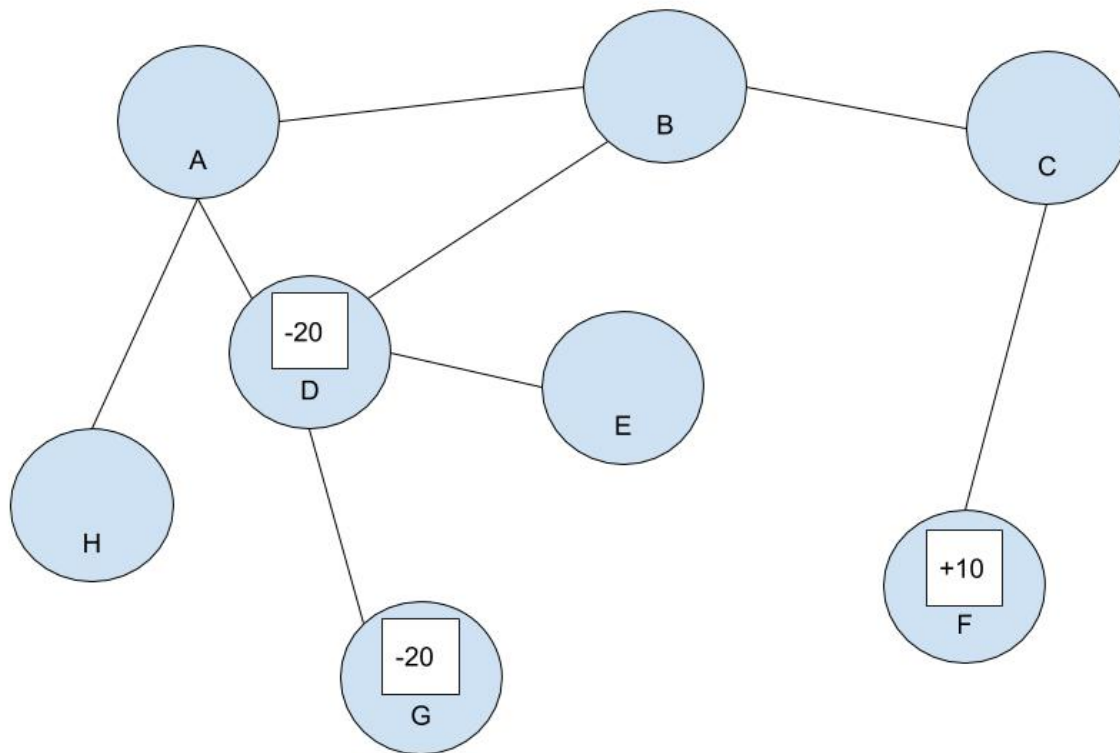
Problem 4: Loaded-dice game tree

Consider a game with the game tree and terminal utilities shown in the picture. First the MAX player makes a move that can be LEFT or RIGHT. Then follows a chance node, with the probabilities indicated in the figure. Finally, the MIN player makes her move of LEFT or RIGHT.



- List the expected utilities at nodes A ... G, in the form $A = 99.3$, $B = \dots$
- Assuming that both MAX and MIN are rational, describe what they are going to play in this game. Use the format: "MAX at A plays LEFT".
- Now assume that MIN, instead of always minimizing, plays as follows: with 80% probability it chooses the minimizing move; with 20% probability it chooses randomly. Calculate the expected utilities at A ... G under this assumption. Explain your reasoning.

Problem 5: Optimal MDP values and policy

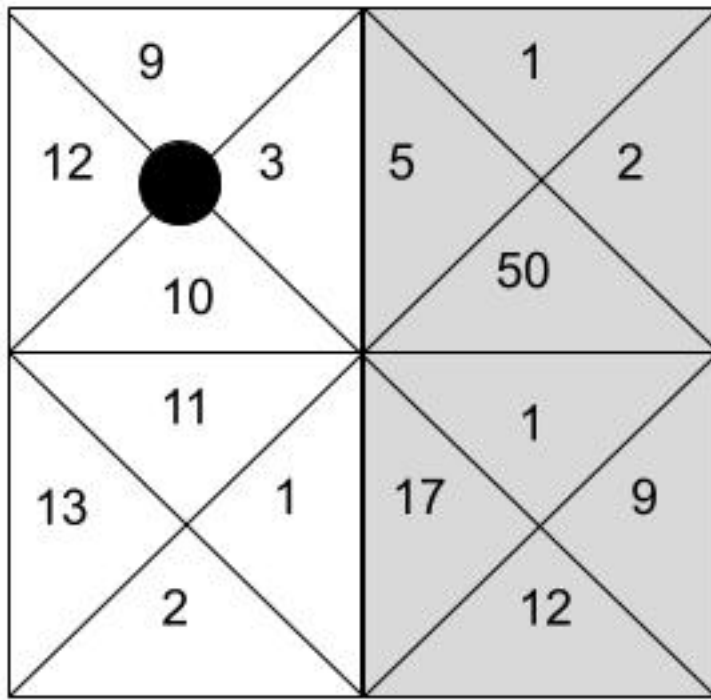


Consider the world described by the graph above. An agent can move between two states only if they are connected (it can move from A to B and from B to A, but not from B to E). The states with square boxes are terminal states, where the only possible action is exit, with the reward written in the square. Every other action has a reward of -2 . Every action always succeeds. Assume $\gamma = 1$.

We will use the following notation:

- $V(X)$: the V-value of state X (e.g., $V(F) = 10$).
 - $Q(X, Y)$: the Q-value of being in state X and taking the action of moving to Y (e.g., $Q(S, P) = -234$). Similarly, $Q(\text{state}, \text{exit})$ is the Q-value of the exit action, e.g., $Q(F, \text{exit}) = 10$.
 - $\pi(X) = Y$: the policy that from X we move to Y.
- a) Using the definition of the optimal V-value, calculate the V-values and list them. Start with a one-sentence explanation of how you are calculating them.
 - b) Using the definition of the optimal Q-value, calculate it for all possible states and actions and list the values. Start with a one-sentence explanation of how you are calculating them.
 - c) For all states, list the optimal policy π . Start with a one-sentence explanation of how you are calculating it.

Problem 6: BlackDisk Q-learning



BlackDisk lives in the world of four squares shown above. Every timestep it must make an action of moving Up, Down, Left, or Right. There is no uncertainty in the movements: BlackDisk always ends up where it is headed, except when it hits the wall, in which case it stays in place. The figure shows its current Q-values. Every time BlackDisk finishes an action in a gray state, it gets a reward of +30. Every time it ends up in a white state, it gets a reward of 0. Every time it hits a wall, it gets a reward of -50.

BlackDisk is trying to improve its life using Q-learning with a learning rate $\alpha = 0.3$ and discount factor $\gamma = 0.5$. Assume that BlackDisk has not yet learned about exploration strategies: it performs exploitation only, following the policy implied by the Q-values.

- 1) Trace two steps of Q-learning. Show the location of the agent after these steps. Show the changes of the Q-values at each step.
- 2) What do you think the optimal policy is for BlackDisk?
- 3) Do you see anything unusual about the initial Q-values? Speculate about where they might have come from.

Problem 7: Behavior cloning versus inverse RL

1. Explain the difference between behavior cloning and inverse reinforcement learning in natural language (do not use formulas). (3 points)
2. It is often claimed that inverse reinforcement learning can create a policy that outperforms the source policy. Explain the reasons for this. (2 points)

Problem 8: Societal challenge and mitigation

1. In one paragraph, state the biggest challenge to society posed by artificial intelligence over the next 5 years. You can think either in terms of the societal system of the USA or in international/global terms, but clearly state which one you are considering. Assume ongoing progress in AI, but no sudden leap in capability. (2 points)
2. In two paragraphs, describe ways in which society might mitigate this challenge. Depending on whether you took the national or international perspective in the previous question, think about the societal actions in terms of the political reality of that setting. (3 points)

Part II: Solutions

Solution 1: Joint probability queries

1. Sum the rows in which allergy is present, marginalizing out the other three variables:

$$0.252 + 0.028 + 0.063 + 0.007 + 0.1944 + 0.0216 + 0.0486 + 0.0054 = 0.62$$

2. Select the rows with sneezing. Fever and sinus pain are hidden variables; marginalize them out:

$$\text{no allergy: } 0.0216 + 0.0024 + 0.0054 + 0.0006 = 0.03$$

$$\text{allergy: } 0.1944 + 0.0216 + 0.0486 + 0.0054 = 0.27$$

Normalize by dividing by 0.30: $P(\text{allergy} \mid \text{sneezing}) = 0.27 / 0.30 = 0.9$.

3. Select the rows with fever and not sneezing. Sinus pain is hidden; marginalize it out:

$$\text{no allergy: } 0.063 + 0.007 = 0.070$$

$$\text{allergy: } 0.063 + 0.007 = 0.070$$

Normalizing gives $P(\text{allergy} \mid \text{fever, not sneezing}) = 0.5$.

4. Yes. Marginalize out sneezing and allergy to obtain the joint distribution of fever and pain:

$$\text{no fever, no pain: } 0.252 + 0.252 + 0.0216 + 0.1944 = 0.72$$

$$\text{fever, no pain: } 0.063 + 0.063 + 0.0054 + 0.0486 = 0.18$$

$$\text{no fever, pain: } 0.028 + 0.028 + 0.0024 + 0.0216 = 0.08$$

$$\text{fever, pain: } 0.007 + 0.007 + 0.0006 + 0.0054 = 0.02$$

Hence $P(\text{fever}) = 0.2$ and $P(\text{pain}) = 0.1$. Independence requires $P(\text{Pain, Fever}) = P(\text{Pain}) P(\text{Fever})$ for every combination of values. Here $0.02 = 0.1 \times 0.2$, and the other three cells also match ($0.18 = 0.9 \times 0.2$, $0.08 = 0.1 \times 0.8$, $0.72 = 0.9 \times 0.8$). An equivalent test is $P(\text{pain} \mid \text{fever}) = 0.02 / 0.2 = 0.1 = P(\text{pain})$.

Points. Question 1: 2 points. Questions 2, 3, and 4: 1 point each. Total: 5 points.

Solution 2: Search algorithm comparison

- **1-Left:** finds a solution, and it is optimal. Not very efficient.
- **2-Left:** finds a solution, and it is optimal. Not very efficient. No difference from breadth-first tree search, because the Left graph has no repeated states.
- **3-Left:** finds a solution, and it is optimal. In this case it is very efficient.
- **4-Left:** finds the solution D^* very efficiently.
- **5-Left:** guaranteed to find the optimal solution, and it does. The Euclidean distance is admissible and consistent here, because it is a relaxation of the problem.
- **1-Right:** finds a solution, and it is optimal.
- **2-Right:** finds a solution, and it is optimal. It saves no expansions compared to tree search: the second copy of the node above S would be expanded only after the goal is found.
- **3-Right:** finds a solution, but not the optimal one. It follows $S \rightarrow \text{above-S} \rightarrow \text{right of S} \rightarrow \text{below S} \rightarrow \text{left of S}$, and there the upper D (about 1 o'clock) comes before the node above S (1:30). It reaches the upper D in 5 steps, while the optimal solution takes 2. It escapes the cycle only because of the expansion order; with the cycle edge first, it would loop forever.
- **4-Right:** finds a solution, but expands many nodes first. Depending on how the tie between the two copies of the node left of S is broken, it may not find the optimal solution.

- **5-Right:** not guaranteed to find the optimal solution. The heuristic is neither admissible nor consistent: the node to the left of S has a true distance of 1 to D*, but a heuristic of 3. On this particular instance, however, A* does find the optimal solution. It expands S, the node above S ($f = 1 + \sqrt{5} \approx 3.24$), and the node right of S ($f = 2 + 1 = 3$). Then the node left of S ($f = 1 + 3 = 4$) has the lowest f-value, and its successor D* ($f = 2$) is the goal. The path is $S \rightarrow \text{left of } S \rightarrow D^*$, with cost 2.

Points. Each of the 10 answers: 0.5 points. Total: 5 points.

Solution 3: Connect Four evaluation function

- At the beginning of the game the branching factor is 7 (in general, n , the number of columns). It decreases as columns fill up.
- Several choices are possible, for example:
 - the number of extendable-on-both-sides 3-connections of the player minus the number of extendable 3-connections of the opponent;
 - as above, but also counting 2-connections;
 - as above, distinguishing connections extendable on both sides from those extendable on one side, with the former weighted more.
- The answer depends on the function chosen. The student needs to reflect on the cost of evaluating it. For instance, the first function above requires at least $O(mn)$ work.
- The student has to argue that a higher value makes it more likely that the player wins. For instance, a player who has the turn and an extendable 3-connection can win.
- Most evaluation functions are not perfect. The student should point out one of the following:
 - the function gives no useful value early enough (for instance, the 3-connection count is zero until extendable 3-connections exist);
 - the function misses subtleties (for instance, three extendable 2-connections might all be blocked by a single piece);
 - a perfect evaluation function is equivalent to solving the game to the end.

Points. Each of the five questions: 1 point. Total: 5 points.

Solution 4: Loaded-dice game tree

- $A = -3.6, B = -3.6, C = -5.8, D = -5, E = -3, F = -10, G = -4$.

$$D = \min(3, -5) = -5 \quad E = \min(2, -3) = -3$$

$$F = \min(-10, 4) = -10 \quad G = \min(5, -4) = -4$$

$$B = 0.3 \times -5 + 0.7 \times -3 = -3.6$$

$$C = 0.3 \times -10 + 0.7 \times -4 = -5.8$$

$$A = \max(-3.6, -5.8) = -3.6$$

(b)

MAX at A plays LEFT

MIN at D plays RIGHT

MIN at E plays RIGHT

MIN at F plays LEFT

MIN at G plays RIGHT

(c) A random choice puts probability 0.1 on each child, so a MIN node is worth $0.8 \times \text{min} + 0.1 \times \text{left} + 0.1 \times \text{right}$:

$$\begin{aligned} D &= 0.8 \times -5 + 0.1 \times 3 + 0.1 \times -5 = -4.2 \\ E &= 0.8 \times -3 + 0.1 \times 2 + 0.1 \times -3 = -2.5 \\ F &= 0.8 \times -10 + 0.1 \times -10 + 0.1 \times 4 = -8.6 \\ G &= 0.8 \times -4 + 0.1 \times 5 + 0.1 \times -4 = -3.1 \\ B &= 0.3 \times -4.2 + 0.7 \times -2.5 = -3.01 \\ C &= 0.3 \times -8.6 + 0.7 \times -3.1 = -4.75 \\ A &= \max(-3.01, -4.75) = -3.01 \end{aligned}$$

Points. (a) 2 points, (b) 2 points, (c) 1 point. Total: 5 points.

Solution 5: Optimal MDP values and policy

a) For terminal states, V is the exit reward. For the others, $V(X) = \max$ over neighbors Y of $[-2 + \gamma V(Y)]$, computed backward from the terminal states:

$$\begin{aligned} V(F) &= 10, \quad V(D) = -20, \quad V(G) = -20 \\ V(C) &= -2 + V(F) = 8 \\ V(B) &= -2 + V(C) = 6 \\ V(A) &= -2 + V(B) = 4 \\ V(H) &= -2 + V(A) = 2 \\ V(E) &= -2 + V(D) = -22 \end{aligned}$$

b) $Q(X, Y) = -2 + \gamma V(Y)$ for moves; $Q(X, \text{exit})$ is the exit reward:

$$\begin{aligned} Q(A, B) &= 4 & Q(A, H) &= 0 & Q(A, D) &= -22 \\ Q(B, C) &= 6 & Q(B, A) &= 2 & Q(B, D) &= -22 \\ Q(C, F) &= 8 & Q(C, B) &= 4 & & \\ Q(H, A) &= 2 & & & & \\ Q(E, D) &= -22 & & & & \\ Q(D, \text{exit}) &= -20 & Q(G, \text{exit}) &= -20 & Q(F, \text{exit}) &= 10 \end{aligned}$$

c) $\pi(X) = \text{argmax over actions of } Q(X, \cdot)$:

$$\begin{aligned} \pi(A) &= B, \quad \pi(B) = C, \quad \pi(C) = F, \quad \pi(H) = A, \quad \pi(E) = D, \\ \pi(D) &= \text{exit}, \quad \pi(F) = \text{exit}, \quad \pi(G) = \text{exit} \end{aligned}$$

Points. a) 2 points, b) 2 points, c) 1 point. Total: 5 points.

Solution 6: BlackDisk Q-learning

The update rule:

$$\text{sample} = R(s, a, s') + \gamma \max_{a'} Q(s', a')$$

$$Q(s, a) \leftarrow (1 - \alpha) Q(s, a) + \alpha \cdot \text{sample}$$

- 1) **Step 1.** In top-left, Left has the largest Q-value (12). BlackDisk hits the wall and stays in top-left; reward -50. $s = s' = \text{top-left}$.

$$\text{sample} = -50 + 0.5 \times 12 = -44$$

$$Q(\text{top-left}, \text{Left}) \leftarrow 0.7 \times 12 + 0.3 \times (-44) = -4.8$$

All other values are unchanged.

Step 2. In top-left, Down now has the largest Q-value (10). BlackDisk moves to bottom-left; reward 0. $s = \text{top-left}$, $s' = \text{bottom-left}$.

$$\text{sample} = 0 + 0.5 \times 13 = 6.5$$

$$Q(\text{top-left}, \text{Down}) \leftarrow 0.7 \times 10 + 0.3 \times 6.5 = 8.95$$

After two steps BlackDisk is in bottom-left.

- 2) From a white state, move into the adjacent gray state (Right). In a gray state, move to the other gray state (Down from top-right, Up from bottom-right). For reference, with $\gamma = 0.5$ every state has $V^* = 60$.
- 3) Both white states have their largest Q-value on a wall collision (Left = 12 in top-left, Left = 13 in bottom-left), although a collision costs -50. The bottom-right gray state also prefers moving Left into a white state over moving Up into the other gray state. These values are inconsistent with the converged Q-values for this reward structure. For comparison, Q^* is 60 for moves into a gray state, 30 for moves into a white state, and -20 for a white-state wall collision. Plausible origins include arbitrary or random initialization, early learning with little exploration, a different earlier reward structure (e.g., no wall penalty), or transfer from another environment.

Points. Question 1: 2 points. Question 2: 2 points. Question 3: 1 point. Total: 5 points.

Solution 7: Behavior cloning versus inverse RL

1. Behavior cloning treats the demonstrations as supervised learning data and directly learns a policy from them. Inverse RL learns a reward function from the demonstration data, then uses it to define an MDP from which a policy is obtained (by planning or RL).
2. Even unsuccessful or suboptimal demonstrations can reveal what the demonstrator was trying to achieve. IRL recovers a reward function that captures this goal; optimizing that reward can then yield a policy that pursues the goal more effectively than the demonstrator did.

Points. Question 1: 3 points. Question 2: 2 points. Total: 5 points.

Solution 8: Societal challenge and mitigation

There is no single correct answer. A complete answer to question 1 names one specific challenge, states the chosen framing, and explains why the challenge is significant within 5 years without assuming a capability leap. A complete answer to question 2 proposes concrete mitigations and explains why they are feasible given the political reality of the chosen setting.

Points. Question 1: 2 points. Question 2: 3 points. Total: 5 points.