University of Illinois at Urbana-Champaign

CS440/ECE448 Artificial Intelligence

Conflict Exam 3

Spring 2026

Main Exam is April 6, 2026

Your Name: ______________________________

Your NetID: ______________________________

Instructions

- Please write your name on the top of every page.
- Have your ID ready; you will need to show it when you turn in your exam.
- This will be a CLOSED BOOK, CLOSED NOTES exam. You are permitted to bring and use only one 8.5x11 page of notes, front and back, handwritten.
- No electronic devices (phones, tablets, calculators, computers etc.) are allowed.
- Make sure that your answer includes only the variables that it should include, but DO NOT simplify explicit numerical expressions, including expressions involving standard functions (exp, log, sin, cos) and expressions involving the sum, product, maximum, or argmax of a complete list of alternatives. For example, the answer $x = \max\left(\sin(0.3), \frac{1}{1+\exp(-0.1)}\right)$ is MUCH preferred (much easier for us to grade) than the answer $x = 0.524979$.

Possibly Useful Formulas

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad \sigma'(z) = \sigma(z)\left(1-\sigma(z)\right)$$

$$f = g\left(\mathbf{w}^T\mathbf{x}\right), \quad \frac{\partial f}{\partial \mathbf{w}} = g'\left(\mathbf{w}^T\mathbf{x}\right)\mathbf{x}$$

$$\mathrm{softmax}(\mathbf{z}) = \frac{\exp(\mathbf{z})}{\sum_i \exp(z_i)}, \quad \frac{\partial\, \mathrm{softmax}_k(\mathbf{z})}{\partial \mathbf{z}} = \mathrm{softmax}_k(\mathbf{z})\left(\mathbb{1}_k - \mathrm{softmax}(\mathbf{z})\right)$$

$$f[k] = \sum_i w[i]x[k-i], \quad \frac{\partial f[k]}{\partial w[i]} = x[k-i]$$

$$\mathbf{C} = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V}$$

$$S : \{\mathscr{V}_P, \mathscr{V}_Q\} \rightarrow \{\mathscr{V}_Q, \mathscr{C}\} \text{ s.t. } S(P) = S(Q) = U$$

$$\frac{x'}{x} = \frac{y'}{y} = \frac{-f}{z}$$

$$R_{c,d} = \frac{\frac{\partial z_c}{\partial x_d} x_d}{\sum_{d'} \frac{\partial z_c}{\partial x_{d'}} x_{d'}} R_c$$

$$\mathscr{C}_{\mathrm{obs}} = \{\mathbf{c} : \exists \mathbf{b} : \phi(\mathbf{b},\mathbf{c}) \in \mathscr{W}_{\mathrm{obs}}\}$$

Admissible: $\hat{h}(n) \leq h(n)$

Consistent: $\hat{h}(n) - \hat{h}(m) \leq d(n,m)$

Bellman: $u(s) = r(s) + \gamma \max_a \sum_{s'} P(s'|s,a)u(s')$

Policy Eval: $u_i(s) = r(s) + \gamma \sum_{s'} P(s'|s,\pi_i(s))u_i(s')$

Policy Update: $\pi_{i+1}(s) = \arg\max_a \sum_{s'} P(s'|s,a)u_i(s')$

Value Iteration: $u_{i+1}(s) = r(s) + \gamma \max_a \sum_{s'} P(s'|s,a)u_i(s')$

Model-based: $P(s_{t+1}|s_t,a_t) = \frac{N(s_t,a_t,s_{t+1})+k}{\sum_{s'}\left(N(s_t,a_t,s')+k\right)}$

TD-Learning: $q(s_t,a_t) \leftarrow (1-\eta)q(s_t,a_t) + \eta\left(r_t + \gamma \max_a q_t(s_{t+1},a)\right)$

SARSA: $q(s_t,a_t) \leftarrow (1-\eta)q(s_t,a_t) + \eta\left(r_t + \gamma q_t(s_{t+1},a_{t+1})\right)$

Actor-Critic: $\mathscr{L}_{\mathrm{actor}} = -\sum_a \pi_a(s)q(s,a)$

REINFORCE: $\Delta\mathbf{W} = \eta(r-\mu)\sum_t \frac{\partial \log \pi_{a_t}(s_t)}{\partial \mathbf{W}}$

Question 1 ***(28 points)***

Consider a neural network with the following architecture. The input features are $\mathbf{x} = [x_1, \ldots, x_n]^T$, the layer weights are $\mathbf{w} = [w_1, \ldots, w_n]^T$, and there is a scalar weight u. The scalar outputs f, g, and h are defined as

$$f = \mathbf{w}^T \mathbf{x},$$
$$g = \sigma(f),$$
$$h = ug$$

where $\sigma(\cdot)$ is the logistic sigmoid.

(a) (14 points) Suppose that you know $\frac{\partial \mathscr{L}}{\partial h}$. In terms of $\frac{\partial \mathscr{L}}{\partial h}$ and any of the variables f, g, h, and/or $\mathbf{w}$, find $\frac{\partial \mathscr{L}}{\partial g}$.

Solution:

$$\frac{\partial \mathscr{L}}{\partial g} = u \frac{\partial \mathscr{L}}{\partial h}$$

(b) (14 points) Suppose that you know $\frac{\partial \mathscr{L}}{\partial g}$. In terms of $\frac{\partial \mathscr{L}}{\partial g}$, and in terms of f, g, h, and/or $\mathbf{x}$, find $\frac{\partial \mathscr{L}}{\partial \mathbf{w}}$.

Solution:

$$\frac{\partial \mathscr{L}}{\partial \mathbf{w}} = \frac{\partial \mathscr{L}}{\partial g} \frac{\partial g}{\partial f} \frac{\partial f}{\partial \mathbf{w}} = \mathbf{x} \sigma(f)(1 - \sigma(f)) \frac{\partial \mathscr{L}}{\partial g}$$

The answer $\mathbf{x} g(1-g) \frac{\partial \mathscr{L}}{\partial g}$ is also correct.

Question 2 ***(15 points)***

Consider a one-dimensional convolutional neural network with input image $x[i,j]$, weights $w[i]$, and output image $f[i,j]$ related as:

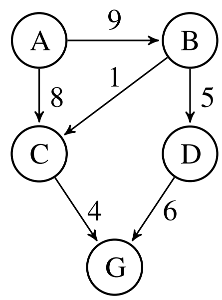
$$f[k,l] = \sum_j x[k,j]w[l-j]$$

In terms of the elements of x, w, and/or f, what is $\frac{\partial f[k,l]}{\partial x[i,j]}$?

Solution:

$$\frac{\partial f[k,l]}{\partial x[i,j]} = \begin{cases} w[l-j] & k=i \\ 0 & \text{otherwise} \end{cases}$$

Question 3 *(28 points)*



(a) (14 points) The search graph above starts at node A, and ends at node G. Create a table showing, in the first column, the node that is expanded by breadth-first search at each step of the search process (starting with A, ending with G), and in the second column, the set of nodes that are in the frontier after the node in the first column has been expanded. Optionally, you may list a priority next to each node in the frontier if you wish. Note that, since BFS is complete, you do not need to put a node on the frontier a second time if it has already been placed there. Ties are broken in alphabetical order.

Solution:

Expand	Frontier
A	B:1, C:1
B	C:1, D:2
C	D:2, G:3
D	G:3
G	

(b) (14 points) Suppose you're given the following partial heuristic for an A* search: $\hat{h}(C) = 2, \hat{h}(D) = 2, \hat{h}(G) = 0$. You must now choose the heuristics $\hat{h}(A)$ and $\hat{h}(B)$. What is the definition of a **consistent** heuristic? What are the largest values of $\hat{h}(A)$ and $\hat{h}(B)$ that would result in a consistent heuristic?

Solution: A consistent heuristic is one for which $\hat{h}(n) - \hat{h}(m) \leq h(n,m)$, the distance from n to m. The largest values that provide a consistent heuristic are $\hat{h}(A) = 10, \hat{h}(B) = 3$.

Question 4 ***(29 points)***

Consider an MDP with two states, $s \in \{0,1\}$, and two actions, $a \in \{0,1\}$. The states have rewards $r(0)=3$ and $r(1)=7$, and the transition probabilities are:

s,a	$P(s'=0\|s,a)$	$P(s'=1\|s,a)$
$0,0$	0.8	0.2
$0,1$	0.3	0.7
$1,0$	0.4	0.6
$1,1$	0.1	0.9

(a) (15 points) Suppose you're performing policy iteration, and with an initial policy of $\pi_1(s)=1$ (i.e., always perform action 1), you learn that the resulting utilities are $u_1(0)=9$ and $u_1(1)=14$. The next step is policy update, which chooses a new policy, $\pi_2(s)$. What are $\pi_2(0)$, the new policy for state $s=0$?

Solution:

$$\pi_2(s) = \arg\max_a \sum_{s'} P(s'|s,a)u_1(s')$$

$$\pi_2(0) = \arg\max\left(P(0|0,0)u_1(0)+P(1|0,0)u_1(1),\right.$$
$$\left.P(0|0,1)u_1(0)+P(1|0,1)u_1(1)\right)$$
$$= \arg\max\left(0.8\cdot 9+0.2\cdot 14, 0.3\cdot 9+0.7\cdot 14\right)$$

(b) (14 points) Explain in words why most studies of MDP must use a discount factor, γ, that is between 0 and 1.

Solution: If you have an infinite-length path, then even with finite rewards, the sum of all future rewards is infinite, which is bad because if two different paths give infinite rewards, there's no way to say which is the best path. With a discount factor $0<\gamma<1$, the sum of all discounted future rewards is guaranteed to be finite, which allows us to choose the best path.

This page is scratch paper

This page is scratch paper, you may tear it off if you wish