UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CS440/ECE448 Artificial Intelligence

Exam 2

Spring 2023

April 3, 2023

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 or typed in a font size comparable to handwriting.
- No electronic devices (phones, tablets, calculators, computers etc.) are allowed.
- SHOW YOUR WORK. Correct answers derivation may not receive full credit if you don't show your work.
- Make sure that your answer includes only the variables that it should include. Solve integrals and summations. After that is done, do not further simplify explicit numerical expressions. For example, the answer $x = \frac{1}{1+\exp(-0.1)}$ is MUCH preferred (much easier for us to grade) than the answer $x = 0.524979$.

Possibly Useful Formulas

Consistent Heuristic: $h(p) \leq d(p,r) + h(r)$

Alpha-Beta Max Node: $v = \max(v, \text{child}); \quad \alpha = \max(\alpha, \text{child})$

Alpha-Beta Min Node: $v = \min(v, \text{child}); \quad \beta = \min(\beta, \text{child})$

Variance Network: $\mathscr{L} = \frac{1}{n-1}\sum_{i=1}^{n}\left(f_2(x_i) - (f_1(x_i) - x_i)^2\right)^2$

Unification: $U = S(P) = S(Q);\ U \Rightarrow \exists x : Q;\ U \Rightarrow \exists x : P$

Bayes Rule: $P(Y=y|X=x) = \frac{P(X=x|Y=y)P(Y=y)}{\sum_{y'} P(X=x|Y=y')P(Y=y')}$

Unnormalized Relevance: $\tilde{R}(f_c, x_d) = \frac{\partial f_c}{\partial x_d} x_d f_c$

Normalized Relevance: $R(f_c, x_d) = \frac{\frac{\partial f_c}{\partial x_d} x_d}{\sum_{d'} \frac{\partial f_c}{\partial x_{d'}} x_{d'}} f_c$

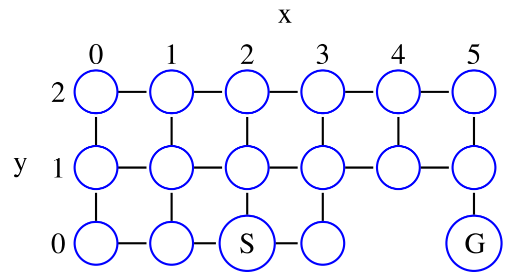
Softmax: $\text{softmax}_j(e) = \frac{\exp(e_j)}{\sum_k \exp(e_k)}$

Softmax Deriv: $\frac{\partial\, \text{softmax}_m(e)}{\partial e_n} = \text{softmax}_m(e)\delta[m-n] - \text{softmax}_m(e)\,\text{softmax}_n(e), \quad \delta[m-n] = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$

Viterbi: $v_t(j) = \max_i v_{t-1}(i) a_{i,j} b_j(x_t)$

Transformer: $c_i = \text{softmax}(q_i @ k^T) @ v$

Question 1 ***(8 points)***

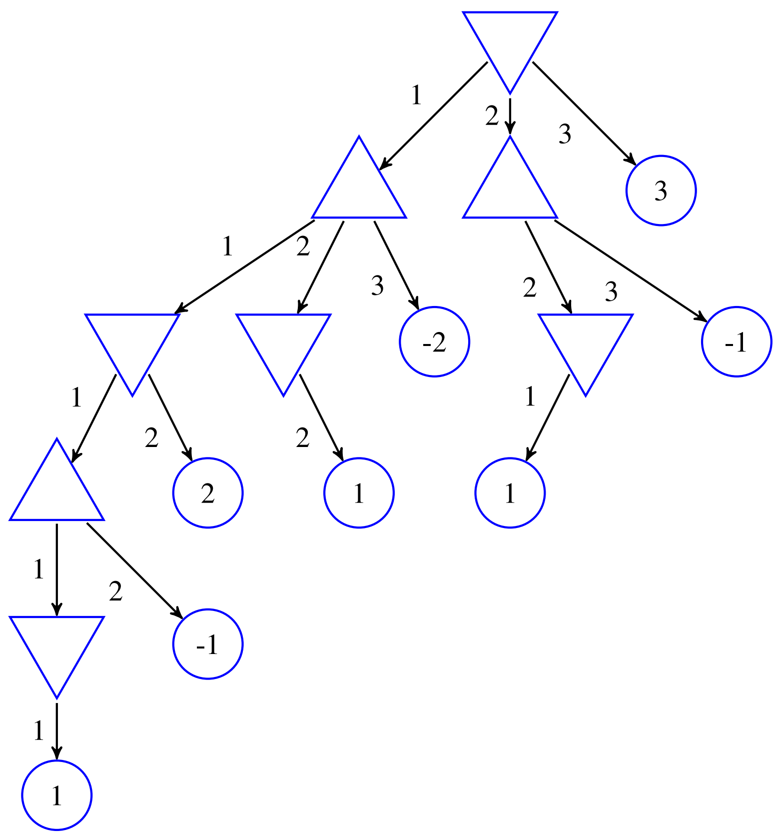


In the maze shown above, nodes are named by their (y,x) coordinates, where y is the row number (starting from the bottom), and x is the column number (starting from the left). A robot is trying to find a path from the start node, $(0,2)$ (labeled "S"), to the goal node, $(0,5)$ (labeled "G"). It uses A* search, with Manhattan distance as a heuristic. After nodes $(0,3)$ and $(1,2)$ have been expanded, there are two copies of nodes $(1,3)$ on the frontier, one with $(0,3)$ as its parent, and one with $(1,2)$ as its parent. Which of these two copies was placed on the frontier first? Why?

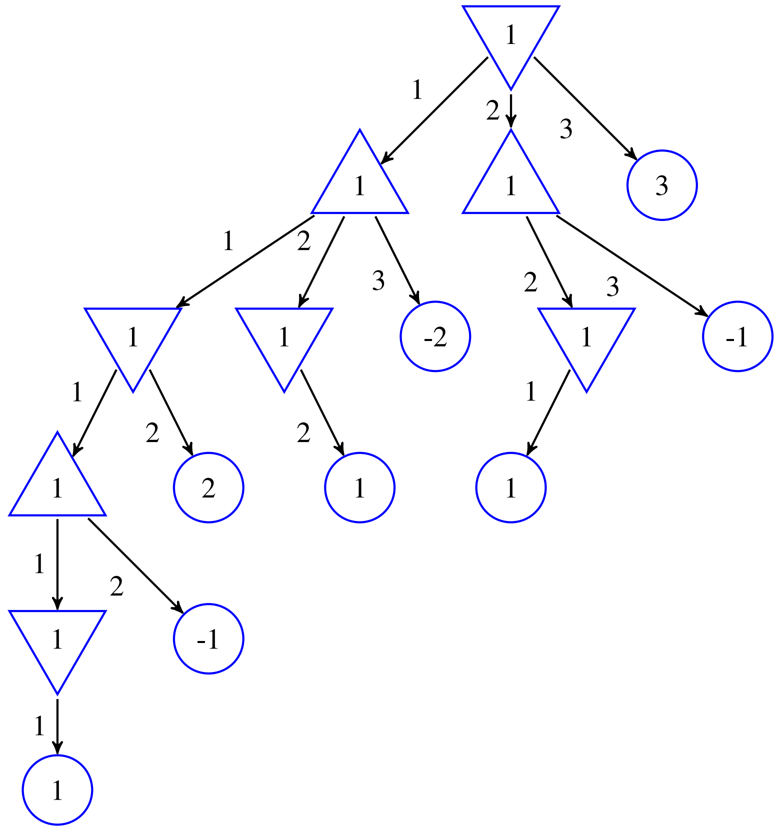
Solution: Both nodes $(0,3)$ and $(1,2)$ are one step away from the start node ($g((0,3)) = g((1,2)) = 1$), but node $(0,3)$ has a lower heuristic ($h((0,3)) = 2$, while $h((1,2)) = 4$), therefore node $(0,3)$ is expanded before node $(1,2)$. The copy of $(1,3)$ with $(0,3)$ as its parent is therefore placed on the frontier before the copy that has $(1,2)$ as its parent.

Question 2 ***(7 points)***

Consider a game in which Max and Min each start with three stones. Min plays first. On their turn, each player must discard one, two, or three stones. The number of stones a player discards must be greater than or equal to the number of stones their opponent discarded on the immediate preceding turn (if a player does not have enough stones to satisfy this rule, they must discard all of their remaining stones). When one player loses their last stone, the other player wins a number of points equal to the number of stones they have not yet discarded. The figure below shows the game tree for this game. In each triangle (each Min or Max node), enter a number specifying the value of that node. Specify the minimax value of the game. Indicate a minimax move sequence (there are several different sequences that a pair of optimal players might play; specify one of those sequences).



Solution:



The minimax value is +1. One sequence of moves with this value: each player plays 1 stone, in turn, until Min plays her last stone. Another possible solution: Min plays 1, Max plays 2, Min plays 2. Another possible solution: Min plays 2, Max plays 2, Min plays her last stone.

Question 3 ***(7 points)***

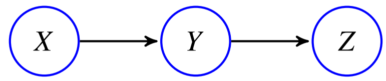
u, v, w, x, y, and z are variables. You are trying to determine whether or not it's possible to perform a step of forward-chaining using the rule $T = \text{Uses}(u, \text{cellphone}) \Rightarrow \text{Human}(u)$. The facts currently available to you in the database $\mathscr{D}$ are:

$$\mathscr{D} = \left\{ \begin{array}{c} P = \text{Eats}(\text{tiger}, \text{cellphone}) \\ Q = \forall v : \text{Uses}(v, \text{landline}) \\ R = \forall w : \text{Uses}(\text{george}, w) \\ S = \exists x : \text{Zambonis}(x, \text{icerink}) \end{array} \right\}$$

Which proposition (P, Q, R, or S) can be unified with the antecedent of T? What is the resulting unified proposition, what is the resulting substitution dictionary, and what new fact is added to the database?

Solution: R can be unified with T, giving $\text{Uses}(\text{george}, \text{cellphone})$, $S = \{u : \text{george}, w : \text{cellphone}\}$, and adding $\text{Human}(\text{george})$ to the database.

Question 4 *(7 points)*



X, Y, and Z are random variables, each of which can take the values -1, 0, or 1. Their causal dependencies are shown in the Bayes network above. The parameters of this model are:

	x		
	-1	0	1
$P(X=x)$	a	b	c
$P(Y=-1\|X=x)$	d	e	f
$P(Y=0\|X=x)$	g	h	i
$P(Y=1\|X=x)$	j	k	l

	y		
	-1	0	1
$P(Z=-1\|Y=y)$	m	n	o
$P(Z=0\|Y=y)$	p	q	r
$P(Z=1\|Y=y)$	s	t	u

In terms of the parameters a through u, what is $P(X=-1|Z=0)$?

Solution:

$$\begin{aligned}
P(X=-1|Z=0) &= \frac{P(X=-1,Z=0)}{P(Z=0)} \\
&= \frac{\sum_y P(X=-1,Y=y,Z=0)}{\sum_y\sum_x P(X=x,Y=y,Z=0)} \\
&= \frac{a(dp+gq+jr)}{a(dp+gq+jr)+b(ep+hq+kr)+c(fp+iq+lr)}
\end{aligned}$$

Question 5 ***(7 points)***

Suppose $f_j = \text{softmax}_j(w@h)$, where w is a 2×2 matrix, and h is a 2-vector. In terms of h_m, $w_{n,m}$, and/or f_k for appropriate values of k, m, and n, what is the unnormalized relevance of h_0 to the output f_1?

Solution:

$$\tilde{R}(f_1, h_0) = \frac{\partial f_1}{\partial h_0} h_0 f_1$$

Let's define $e = w@h$, then we can write:

$$\begin{aligned}
\frac{\partial f_1}{\partial h_0} &= \sum_j \frac{\partial\, \text{softmax}_5(e)}{\partial e_j} \frac{\partial e_j}{\partial h_0} \\
&= \underset{1}{\text{softmax}}(e)(1 - \underset{1}{\text{softmax}}(e)) w_{1,0} - \underset{1}{\text{softmax}}(e)\, \underset{0}{\text{softmax}}(e) w_{0,0} \\
&= f_1(1 - f_1) w_{1,0} - f_1 f_0 w_{0,0}
\end{aligned}$$

Therefore

$$\tilde{R}(f_1, h_0) = f_1^2 h_0((1 - f_1) w_{1,0} - f_0 w_{0,0})$$

Question 6 ***(7 points)***

Slarti is in Paris, attempting to walk home from a pizza restaurant. He is on a sidewalk whose west edge is a steep drop into the river; he wants to make sure he does not fall into the river. Let Y_t be the true distance between Slarti and the cliff edge at time t, measured in meters (m). Suppose you know that $Y_0 = 3$m, for sure. Since Slarti is too full to walk straight, he wobbles as he walks. Since it's foggy, he does not always see clearly: X_t is how far away the cliff edge looks, at time t, which may or may not be equal to the true distance Y_t. The transition probabilities and observation probabilities are

$$P(Y_t = k|Y_{t-1} = j) = \begin{cases} \frac{1}{4} & k = j-1 \\ \frac{1}{2} & k = j \\ \frac{1}{4} & k = j+1 \end{cases}, \quad P(X_t = k|Y_t = j) = \begin{cases} \frac{1}{3} & k = j-1 \\ \frac{1}{3} & k = j \\ \frac{1}{3} & k = j+1 \end{cases}$$

What is $P(Y_2 = 2, X_2 = 2)$?

Solution:

$$\begin{aligned} P(Y_2 = 1, X_2 = 2) &= \sum_{y_1} P(Y_0 = 3, Y_1 = y_1, Y_2 = 2, X_2 = 2) \\ &= P(Y_0 = 3, Y_1 = 2, Y_2 = 2, X_2 = 2) + P(Y_0 = 3, Y_1 = 3, Y_2 = 2, X_2 = 2) \\ &= \left(\frac{1}{2}\right)\left(\frac{1}{4}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) \end{aligned}$$

Question 7 ***(7 points)***

Suppose the the input to a transformer is the sequence of scalar values $v_t = \left(\frac{t}{1000}\right)$, where $0 \leq t \leq 999$. You are trying to find the context, c_i, for a query, q_i, whose inner product with the keys is

$$q_i @ k_t = \begin{cases} 0 & t \in \{500, 501, 502\} \\ -\infty & \text{otherwise} \end{cases}$$

Find the numerical value of c_i.

Solution:

$$c_i = \text{softmax}(q_i @ k^T) @ v$$

$$\text{softmax}(q_i @ k^T) = \left\{ \begin{array}{ll} \frac{1}{3} & t \in \{500, 501, 502\} \\ 0 & \text{otherwise} \end{array} \right.$$

$$\begin{aligned} c_i &= \frac{1}{3}(v_{500} + v_{501} + v_{502} \\ &= \frac{1}{3}\left(\frac{500}{1000} + \frac{501}{1000} + \frac{502}{1000}\right) \end{aligned}$$

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

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