CS 188 Summer 2023 Introduction to Artificial Intelligence Final

- You have 170 minutes.
- The exam is closed book, no calculator, and closed notes, other than two double-sided cheat sheets that you may reference.
- For multiple choice questions,
  - □ means mark **all options** that apply
  - ○ means mark a **single choice**
- For numerical calculation questions, you may leave your answer unsimplified but **show your work**

First name	
Last name	
SID	
Exam Room	
Name and SID of person to the right	
Name and SID of person to the left	
Discussion TAs (or None)	

Honor code: “As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.”

By signing below, I affirm that all work on this exam is my own work, and honestly reflects my own understanding of the course material. I have not referenced any outside materials (other than two double-sided crib sheet), nor collaborated with any other human being on this exam. I understand that if the exam proctor catches me cheating on the exam, that I may face the penalty of an automatic "F" grade in this class and a referral to the Center for Student Conduct.

Signature: ______________________________

Point Distribution

Q1.	Potpourri: Blast from the Past	10
Q2.	Bayes Nets: Across the Spider-Verse	13
Q3.	HMMs: Head TA Kirby	12
Q4.	Decision Networks: Coin Toss Game	15
Q5.	MDPs: Jim & Pam Part 2	13
Q6.	Reinforcement Learning: Island Hopping	12
Q7.	Machine Learning: Hotdog vs. Not Hotdog	13
Q8.	Neural Nets: Quadratic Activation Function	10
	Total	98

Q1. [10 pts] Potpourri: Blast from the Past

(a) Search. Barbie and Ken are in a building of size $H \times W \times L$. Each cell of the building may or may not contain a toy, and their mission is to collect all of the toys. It is okay for Barbie and Ken to be in the same cell of the building.

(i) [2 pts] Propose a minimal state space representation for this problem.

Current locations of Barbie and Ken in the building, and a Boolean variable indicating the presence of toys for each square.

(ii) [2 pts] What is the size of the state space?

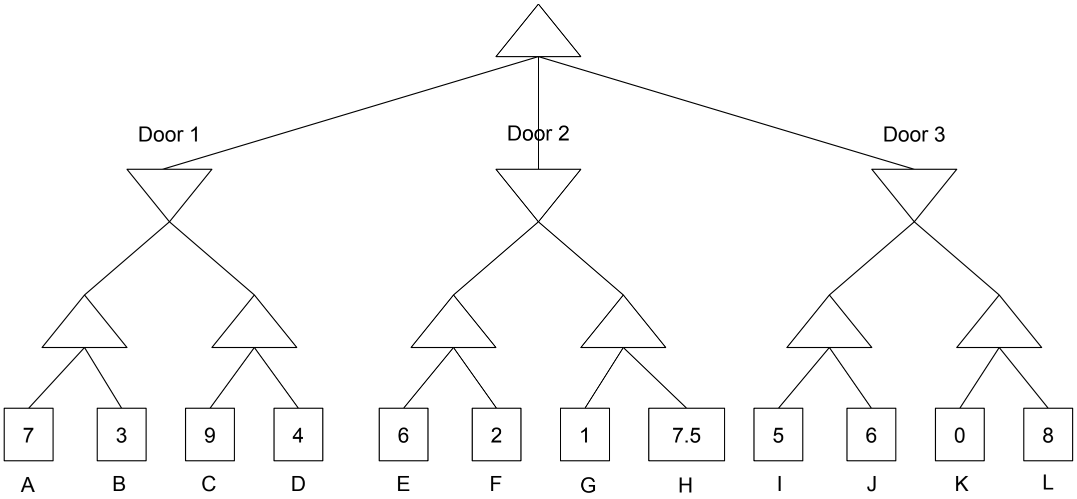
$(HWL)^2 \cdot 2^{HWL}$

(b) [2 pts] **CSPs.**

Alice is scheduling job interviews. Nine different companies reached out to interview her this coming week, and she is panicked trying to schedule all of them in just five days! For the nine companies, three are big (B_1, B_2, B_3), three are medium (M_1, M_2, M_3), and three are small (S_1, S_2, S_3). Write down her constraints formally below:

Index	Explanation	Constraint
A	You should interview with B_2 on Friday.	$B_2 = 5$
B	You should interview with B_3 on Monday.	$B_3 = 1$
C	You should interview with S_1 on either Monday or Tuesday.	$S_1 \in \{1, 2\}$
E	You should interview with S_2 after S_1 (cannot be on the same day).	$S_2 > S_1$
G	You should take at least two days break after M_2 before M_3 (If M_2 occurs on Monday, the earliest M_3 can occur is Thursday).	$M_3 > M_2 + 2$
H	You should interview with M_3 after B_1 (cannot be on the same day), since they have the same interview style.	$M_3 > B_1$

(c) [4 pts] **Games.**



(i) [2 pts] Fill out the values on the game tree. Which door does the top maximizer node choose?

● Door 1 ○ Door 2 ○ Door 3

(ii) [2 pts] Which terminal nodes are never explored as a consequence of pruning? (Assume that we prune on equality.)

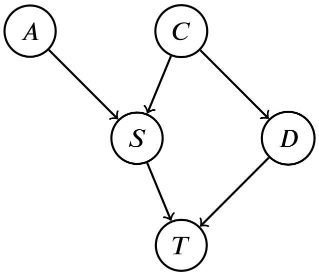
□ A □ B □ C ■ D □ E □ F ■ G ■ H □ I □ J ■ K ■ L

○ None of the above

Q2. [13 pts] Bayesian Networks: Across the Spider-Verse

Miles is curious about the probability that he arrives to school on time. The factors involved with him arriving to school on time can be represented by the following Bayes Net (assume that each variable is a binary variable):

- A: Sets alarm
- S: Over sleeps
- D: Dad is late
- C: Fighting crime last night
- T: Arrives at school on time



(a) [7 pts] Miles wants to calculate $P(C|T)$ using variable elimination. Assume he eliminates variables in alphabetical order (A, D, S).

(i) [1 pt] What factors does he have available at the start?

$P(A), P(C), P(S|A,C), P(D|C), P(T|S,D)$

(ii) [1 pt] First, he eliminates A, and get the new factor

$f_1(C,S) = \boxed{\sum_A P(A) * P(S|A,C)}$

Write out the remaining factors

$P(C), P(D|C), P(T|S,D), f_1(C,S)$

(iii) [1 pt] Then, he eliminates D, and get the new factor

$f_2(C,S,T) = \boxed{\sum_D P(D|C) * P(T|S,D)}$

Write out the remaining factors

$P(C), f_1(C,S), f_2(C,S,T)$

(iv) [1 pt] Then, he eliminates S, and get the new factor

$f_3(C,T) = \boxed{\sum_S f_1(C,S) * f_2(C,S,T)}$

Write out the remaining factors

$P(C), f_3(C,T)$

(v) [1 pt] Finally, join any remaining factors to calculate

$f_4(C,T) = \boxed{P(C) * f_3(C,T)}$

(vi) [1 pt] How can he use this to calculate $P(C=+c|T=-t)$? Your answer should be in terms of f_4.

$P(C=+c|T=-t) = \boxed{\frac{f_4(+c,-t)}{f_4(+c,-t)+f_4(-c,-t)}}$

(vii) [1 pt] Order factors f_1, f_2, and f_3 in increasing order of size. $\boxed{f_1} < \boxed{f_3} < \boxed{f_2}$

The following CPTs correspond to the Bayes Net above:

A	C	S	$P(S\|A,C)$
$+a$	$+c$	$+s$	$11/16$
$+a$	$+c$	$-s$	$5/16$
$+a$	$-c$	$+s$	$1/8$
$+a$	$-c$	$-s$	$7/8$
$-a$	$+c$	$+s$	$9/10$
$-a$	$+c$	$-s$	$1/10$
$-a$	$-c$	$+s$	$1/5$
$-a$	$-c$	$-s$	$4/5$

T	D	S	$P(T\|S,D)$
$+t$	$+d$	$+s$	$1/20$
$+t$	$+d$	$-s$	$2/5$
$+t$	$-d$	$+s$	$1/5$
$+t$	$-d$	$-s$	1
$-t$	$+d$	$+s$	$19/20$
$-t$	$+d$	$-s$	$3/5$
$-t$	$-d$	$+s$	$4/5$
$-t$	$-d$	$-s$	0

C	D	$P(D\|C)$
$+c$	$+d$	$1/4$
$+c$	$-d$	$3/4$
$-c$	$+d$	0
$-c$	$-d$	1

C	$P(A)$
$-c$	$4/5$
$+c$	$1/5$

A	$P(A)$
$-a$	$1/4$
$+a$	$3/4$

Miles is now interested in calculating $P(C=+c|T=-t)$ via sampling. He generates the following random samples (assume the variables were generated from left to right):

Sample	A	C	S	D	T
1	$+a$	$-c$	$-s$	$-d$	$+t$
2	$+a$	$-c$	$+s$	$-d$	$-t$
3	$+a$	$+c$	$+s$	$-d$	$-t$
4	$-a$	$-c$	$-s$	$+d$	$-t$
5	$-a$	$-c$	$+s$	$+d$	$-t$

(b) [3 pts] Assuming Miles uses prior sampling:

(i) [1 pt] Bubble in the samples that Miles uses to calculate the final probability.

□ 1 ■ 2 ■ 3 ■ 4 ■ 5

(ii) [2 pts] What is the probability he calculates via prior sampling?

$P(C=+c|T=-t) =$ $\boxed{1/4}$

(c) [3 pts] Now assuming Miles uses likelihood weighting:

(i) [1 pt] What weight does Miles assign to each sample?

Sample 1: $\boxed{0}$ Sample 2: $\boxed{4/5}$ Sample 3: $\boxed{4/5}$

Sample 4: $\boxed{3/5}$ Sample 5: $\boxed{19/20}$

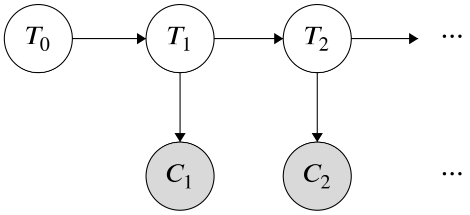
What final probability does he calculate?

(ii) [2 pts] What is the probability he calculates via likelihood weighting?

$P(C=+c|T=-t) =$ $\boxed{16/63}$

Q3. [12 pts] HMMs: Head TA Kirby

Kirby is serving as a TA. He wants to evaluate his teaching performance after each of his weekly discussion sections, and he does so based on how much his students collaborate during his section. He models this situation using an HMM.



T_0	$P(T_0)$
+t	0.6
-t	0.4

T_{i+1}	T_i	$P(T_{i+1}\|T_i)$
+t	+t	0.7
-t	+t	0.3
+t	-t	0.4
-t	-t	0.6

C_i	T_i	$P(C_i\|T_i)$
+c	+t	0.5
-c	+t	0.5
+c	-t	0.1
-c	-t	0.9

T_i is a binary random variable representing whether Kirby taught sufficiently well during Week i. C_i is another binary random variable representing whether students collaborated during Week i. He does not see his students' collaboration during Week 0.

(a) [8 pts] Using the two steps of the forward algorithm, calculate the distribution of $P(T_1|C_1 = +c)$. For the sake of organization, you may use the first box for the **time elapse** update and the second box for the **observation** update. You may also leave your final answers unsimplified (for example, as fractions).

Time elapse update:

$$\begin{aligned} P(T_1 = +t) &= P(T_1 = +t|T_0 = +t)P(T_0 = +t) + P(T_1 = +t|T_0 = -t)P(T_0 = -t) \\ &= 0.7 * 0.6 + 0.4 * 0.4 \\ &= 0.58 \\ P(T_1 = -t) &= P(T_1 = -t|T_0 = +t)P(T_0 = +t) + P(T_1 = -t|T_0 = -t)P(T_0 = -t) \\ &= 0.3 * 0.6 + 0.6 * 0.4 \\ &= 0.42 \end{aligned}$$

Observation update:

$$\begin{aligned}
P(T_1 = +t, C_1 = +c) &= P(C_1 = +c | T_1 = +t) P(T_1 = +t) \\
&= 0.5 * 0.58 = \mathbf{0.29} \\
P(T_1 = -t, C_1 = +c) &= P(C_1 = +c | T_1 = -t) P(T_1 = -t) \\
&= 0.1 * 0.42 = \mathbf{0.042} \\
P(C_1 = +c) &= P(C_1 = +c, C_1 = +c) + P(T_1 = -t, C_1 = +c) \\
&= 0.542 + 0.21 = \mathbf{0.332} \\
P(T_1 = +t | C_1 = +c) &= \frac{P(T_1 = +t, C_1 = +c)}{P(C_1 = +c)} \\
P(T_1 = -t | C_1 = +c) &= \frac{P(T_1 = -t, C_1 = +c)}{P(C_1 = +c)}
\end{aligned}$$

$P(T_1 = +t | C_1 = +c) =$ $\boxed{\frac{0.29}{0.332}}$

$P(T_1 = -t | C_1 = +c) =$ $\boxed{\frac{0.042}{0.332}}$

(b) [4 pts] In order to save computational resources, Kirby turns to particle filtering to analyze this HMM.

(i) [2 pts] At timestep $t = 3$, Kirby has observed the following evidence: $C_1 = +c$, $C_2 = -c$, and $C_3 = +c$. Following the particle filtering algorithm, assign weights to particles in the following states at $t = 3$:

Particles in state $+t$ will have weight: $\boxed{0.5}$

Particles in state -t will have weight: $\boxed{0.1}$

The weight of a particle in some arbitrary state t is $P(C_3 = +c | T_3 = t)$.

(ii) [2 pts] At timestep $t = 6$, we observe 3 particles in state $+t$ and 5 particles in state $-t$, and $C_6 = -c$. Fill in the table describing the distribution that we resample our new particles from for $t = 7$. Show any work in the box on the left.

T_7	$P(T_7)$
+t	0.25
-t	0.75

The 3 particles in state +t each have weight $P(C_6 = -c | T_6 = +t) = 0.5$ (totaling $0.5 * 3 = 1.5$), while the 5 particles in state -t each have weight $P(C_6 = -c | T_6 = -t) = 0.9$ (totaling $0.9 * 5 = 4.5$). Normalizing the weights to form a probability distribution across the domain of T, we arrive at $\mathbf{P(T_7 = +t) = 0.25}$ and $\mathbf{P(T_7 = -t) = 0.75}$.

Q4. [15 pts] Decision Networks: Coin Toss Game

Alice and Bob are participating in a coin toss game. Both players toss two fair coins. Bob's coins are revealed first, after which he must choose to "continue" or "concede". If he concedes, Bob loses \$2 and the game ends. If he continues, Alice's coins are revealed. If Bob's coin toss resulted in strictly more heads than Alice's, he wins \$10. However, if he has an equal or lesser number of heads than Alice, Bob loses \$10. Assume Bob is rational and his utility is the amount of money he wins.

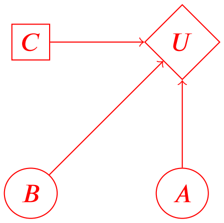
(a) [2 pts] Sketch the decision network for the game, use the following nodes and node types.

Nodes: B represents the number of heads in Bob's coins; A represents the number of heads in Alice's coins; C represents Bob's choice to continue or concede; U represents Bob's utility.

Node types: An elliptical node represents a chance node; a rectangular node represents an action node; a diamond-shaped node represents utility.

C U

B A



(b) [3 pts] For each scenario where Bob gets 0, 1, or 2 heads, what should Bob's decision be: to "continue or "concede"? Justify your answer.

(i) [1 pt] If Bob gets 0 heads:

(ii) [1 pt] If Bob gets 1 heads:

(iii) [1 pt] If Bob gets 2 heads:

By simple probabilities, $\Pr(B = 0) = \Pr(A = 0) = 0.25$, $\Pr(B = 1) = \Pr(A = 1) = 0.5$, and $\Pr(B = 2) = \Pr(A = 2) = 0.25$.

$$\text{EU}(\text{continue}|B = 0) = -10 \tag{1}$$
$$\text{EU}(\text{concede}|B = 0) = -2 \tag{2}$$

So Bob choose to concede when there are 0 heads.

$$\text{EU}(\text{continue}|B = 1) = -10 \cdot 0.75 + 10 \cdot 0.25 = -5 \tag{3}$$
$$\text{EU}(\text{concede}|B = 1) = -2 \tag{4}$$

So Bob choose to concede when there are 1 head.

$$\text{EU}(\text{continue}|B = 2) = -10 \cdot 0.25 + 10 \cdot 0.75 = 5 \tag{5}$$
$$\text{EU}(\text{concede}|B = 2) = -2 \tag{6}$$

So Bob choose to continue when there are 2 heads.

(c) [3 pts] What is the expected monetary gain or loss for Bob in a game?

In the last part, we have $\text{MEU}(B = 0) = -2$ (where Bob concedes), $\text{MEU}(B = 1) = -2$ (where Bob concedes), and $\text{MEU}(B = 2) = 5$ (where Bob chooses to continue).

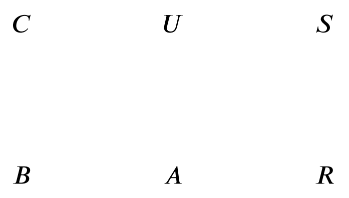
So

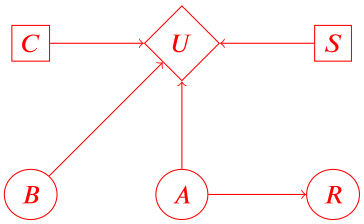
$$\text{EMEU} = \text{MEU}(B = 0)\Pr(B = 0) + \text{MEU}(B = 1)\Pr(B = 1) + \text{MEU}(B = 2)\Pr(B = 2) \tag{7}$$
$$= -2 \cdot 0.25 + (-2) \cdot 0.5 + 5 \cdot 0.25 \tag{8}$$
$$= -0.25 \tag{9}$$

Bob is expected to lose \$0.25 in each game.

(d) [2 pts] Bob perceives the game as unfair and refuses to participate. To persuade him, Alice proposes an additional rule: Bob can pay Alice c dollars ($c > 0$) to see one of Alice's coins *before seeing his own coins*. **Sketch the decision network for the modified coin toss game.**

New node: R represents the revealed coin, either head ($R = H$) or tail ($R = T$). S represents Bob's decision of whether to see one of Alice's coins.





(e) [2 pts] Following the previous part, calculate the conditional distribution of A given R and fill in the table:

R	$\Pr(A = 0 \mid R)$	$\Pr(A = 1 \mid R)$	$\Pr(A = 2 \mid R)$
H			
T			

R	$\Pr(A = 0 \mid R)$	$\Pr(A = 1 \mid R)$	$\Pr(A = 2 \mid R)$
H	0	0.5	0.5
T	0.5	0.5	0

(f) [3 pts] (**Extra Credit!!!!!**, do this last) Following the previous two parts, Alice secretly insists that this new rule should not impact her expected gain. What should the value of c be to meet this condition? Justify your answer.

Alice insists that "this new rule should not affect her expected earnings."

This condition equates to "Bob cannot increase his gain by purchasing the information of R."

This condition further translates to "The cost of the information outweighs its value", i.e., $c \geq \text{EVPI}(R)$ (the expected Value of Perfect Information of R).

Let's calculate the VPI in each case given Bob's number of heads.

If $B = 0$, Bob always loses, and the MEU also remains unchanged. So $\text{VPI}(B = 0) = 0$

If $B = 1$:

$$\text{EU}(\text{continue} \mid B = 1, R = H) = -10 \tag{10}$$
$$\text{EU}(\text{continue} \mid B = 1, R = T) = -10 \cdot 0.5 + 10 \cdot 0.5 = 0 \tag{11}$$
$$\text{EU}(\text{concede} \mid B = 1, R = H) = -2 \tag{12}$$
$$\text{EU}(\text{concede} \mid B = 1, R = T) = -2 \tag{13}$$

Therefore $\text{MEU}(B = 1, R = H) = -2$ (where Bob concedes) and $\text{MEU}(B = 1, R = T) = 0$ (where Bob chooses to continue).

$$\begin{aligned}&\mathrm{VPI}(B=1) && (14)\\ =&(\mathrm{MEU}(B=1,R=H)\Pr(R=H|B=1)+\mathrm{MEU}(B=1,R=T)\Pr(R=T|B=1))-\mathrm{MEU}(B=1) && (15)\\ =&(-2\cdot 0.5+0\cdot 0.5)-(-2) && (16)\\ =&1 && (17)\end{aligned}$$

If $B = 2$:

$$\mathrm{EU}(\text{continue}|B=2,R=H)=-10\cdot 0.5+10\cdot 0.5=0 \tag{18}$$
$$\mathrm{EU}(\text{continue}|B=2,R=T)=10 \tag{19}$$
$$\mathrm{EU}(\text{concede}|B=2,R=H)=-2 \tag{20}$$
$$\mathrm{EU}(\text{concede}|B=2,R=T)=-2 \tag{21}$$

Therefore $\mathrm{MEU}(B = 2, R = H) = 0$ (where Bob chooses to continue) and $\mathrm{MEU}(B = 2, R = T) = 10$ (where Bob chooses to continue).

$$\begin{aligned}&\mathrm{VPI}(B=2) && (22)\\ =&(\mathrm{MEU}(B=2,R=H)\Pr(R=H|B=2)+\mathrm{MEU}(B=2,R=T)\Pr(R=T|B=2))-\mathrm{MEU}(B=2) && (23)\\ =&(0\cdot 0.5+10\cdot 0.5)-5 && (24)\\ =&0 && (25)\end{aligned}$$

The fair price of the information is the expected VPI:

$$\begin{aligned}\mathrm{EVPI}&=\mathrm{VPI}(B=0)\Pr(B=0)+\mathrm{VPI}(B=1)\Pr(B=1)+\mathrm{VPI}(B=2)\Pr(B=2) && (26)\\ &=0.25\cdot 0+0.5\cdot 1+0.25\cdot 0 && (27)\\ &=0.5 && (28)\end{aligned}$$

Therefore, as long as $c \geq 0.5$, Bob cannot increase his gain by purchasing this piece of information, or equivalently, Alice's gain will not be impacted.

Q5. [13 pts] MDPs: Jim & Pam Part 2

Jim and Pam are on a 1x5 grid, where Jim starts at square 1, and Pam is fixed at square 5.

Jim				Pam

In each time step, Jim chooses to either move right or to rest. Choosing to move succeeds with probability p and fails with probability $1-p$, in which case Jim stays in his original square (Jim received 0 utility regardless of success or failure). Choosing to rest always succeeds, and gives $R(d) = 4^{5-d}$ utility, where d is the distance between Jim and Pam. For example, at the start, where $d = 4$, if Jim decides to rest, he gets 4 utility. We represent this as an infinite horizon MDP with no terminal state.

(a) [3 pts] Jim is considering two policies:

Policy 1: Rest at the start forever.

Policy 2: Attempt to move right once, and then, regardless of success or failure, rest forever.

Assuming that Jim starts in square 1, for what values of p is Policy 1 superior to Policy 2 when the discount factor $\gamma = 0.5$? Hint: the sum S of an infinite geometric series with starting value a and ratio r is $S = \frac{a}{1-r}$

Show your work. $0 \le \boxed{0} \le p \le \boxed{1/3} \le 1$

$V_{rest} = 8$ and $V_{move->rest} = 16p + 4(1-p)$, so $8 \geq 16p + 4(1-p)$

(b) [3 pts]

Now assume that $p = 1$. Still assuming that Jim starts in square 1, for what values of γ is Policy 1 superior to Policy 2?

Show your work. $0 \le \boxed{0} < \gamma < \boxed{1/4} \le 1$

$V_{rest} = \frac{4}{1-\gamma}$ and $V_{move->rest} = \frac{16\gamma}{1-\gamma}$, so $\frac{4}{1-\gamma} \geq \frac{16\gamma}{1-\gamma}$

(c) [7 pts] For the following subparts, assume that $\gamma = 0.5$ and $p = 1$.

(i) [3 pts] Perform two iterations of value iteration, for the following locations of Jim. Show your work.

$States$	$s_J = 1$	$s_J = 2$	$s_J = 3$	$s_J = 4$	$s_J = 5$
V_0	0	0	0	0	0
V_1	4	4^2	4^3	4^4	4^5
V_2	$2^{-1} * 4^2$	$2^{-1} * 4^3$	$2^{-1} * 4^4$	$2^{-1} * 4^5$	$4^5 + 2^{-1} * 4^5$

(ii) [3 pts] Perform two iterations of policy iteration for the following locations of Jim.

$States$	$s_J = 1$	$s_J = 2$	$s_J = 3$	$s_J = 4$	$s_J = 5$
π_i	$a_J = right$	$a_J = rest$	$a_J = rest$	$a_J = rest$	$a_J = rest$
V^{π_i}	4^2	$2 * 4^2$	$2 * 4^3$	$2 * 4^4$	$2 * 4^5$
π_{i+1}	$a_J = right$	$a_J = right$	$a_J = right$	$a_J = right$	$a_J = rest$
$V^{\pi_{i+1}}$	$2^{-3} * 4^5$	$2^{-2} * 4^5$	$2^{-1} * 4^5$	$2^0 * 4^5$	$2^1 * 4^5$
π_{i+2}	$a_J = right$	$a_J = right$	$a_J = right$	$a_J = right$	$a_J = rest$

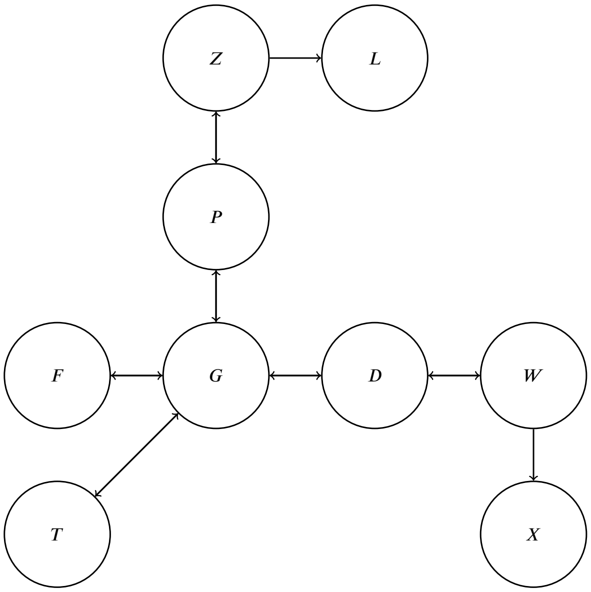
(iii) [1 pt] Assuming that policy iteration has converged, Jim argues it isn't guaranteed values have converged yet, so they need to run value iteration to get the correct values. Pam agrees that policy convergence doesn't guarantee value convergence, but thinks that we don't need to switch to value iteration, as if we continue running policy iteration, eventually the values will converge as well. Who is correct and why?

○ Jim ● Pam

Pam is correct. Policy iteration contains the policy evaluation step, which is essentially just value iteration once the policy has converged.

Q6. [12 pts] Reinforcement Learning: Island Hopping

Bob wants to traverse different islands to reach the island X. He formulates the problem as an MDP where the islands (nodes) represent the states and the arrows represent the possible actions he can take across the seas. Use the direction of the arrows (up, down, left, right) to refer to the specific actions that can be taken.



(a) [2 pts] Bob's ship doesn't always move in the direction that he wants it to. Despite this, he wants to use MLE to build an estimate of the transition function $\hat{T}$ and the reward function $\hat{R}$ for model-based reinforcement learning. He follows some specified policy and collects some data in the form of (current state, action, next state, reward) tuples shown below:

State (s)	Action (a)	New State (s')	Reward
F	Right	G	20
D	Right	G	-10
G	Up	T	-30
P	Down	Z	-15
G	Right	D	30
W	Down	D	-25
G	Right	D	30
D	Left	G	-5
G	Right	T	-30
W	Down	X	100

(i) [1 pt] What is $\hat{T}$(G, Right, D)?

$\frac{2}{3}$

Out of the 3 actions we go right from G, we end up in our desired state of D twice.

(ii) [1 pt] What is $\hat{R}$(W, Down, D)?

-25

(b) [3 pts] Bob looks to use temporal difference learning to learn the values of π, where he has the following initial values:

s	F	G	T	P	Z	L	D	W	X
$V^{\pi}(s)$	5	10	-4	8	2	-10	10	30	50

He performs one update step using the sample (G, Right, D, 30). Assume a discount factor $\gamma = 0.5$ and the learning rate $\alpha = 0.2$. What is the updated value of $V^{\pi}(G)$? Show all your work leading to your answer.

$$V^{\pi}(G) = 15$$

sample $= R(G, Right, D) + 0.5 * V^{\pi}(D) = 30 + 0.5 * 10 = 35$

$V^{\pi}(G) \leftarrow (1 - \alpha) \cdot V^{\pi}(G) + \alpha\cdot$ sample $= 0.8 * 10 + 0.2 * 35 = 15$

(c) [2 pts] Bob wants to have a greedy policy that will minimize exploration and maximize exploitation. Which of the following functions f will do so? Assume k is a positive real number, $N(s,a)$ represents the number of times that the state-action pair is taken, and ϵ is a very small number greater than 0.

- □ $f(s,a) = Q(s,a) + \frac{k}{N(s,a)}$
- ■ $f(s,a) = Q(s,a) + k \cdot e^{N(s,a)}$
- □ $f(s,a) = k \cdot Q(s,a) - log(N(s,a))$
- □ $f(s,a) = \frac{\epsilon}{k \cdot Q(s,a) \cdot N(s,a)}$
- ■ $f(s,a) = \frac{N(s,a) \cdot Q(s,a)}{\epsilon}$
- ○ None of the above

(1) - As the number of state-action pairs increases, the benefit of exploring one particular route converges to zero.

(2) - This value blows up the more times a particular route is explored, since the exponent here is positive.

(3) - Same reasoning as (1), the logarithm term will dominate the first product term as the number of state-action pairs increases.

(4) - The denominator is amplified as the number of times a particular route is explored, so the overall term will converge to zero.

(5) - Same reasoning as (4), but since the product is in the numerator and the denominator is a really small value ϵ, the function will be unbounded and converge to infinity, so one particular route will be preferred.

(d) [5 pts] Bob now switches to Q-learning, where he wants to perform approximate Q-learning for $Q(G, right)$. Assume he has w_i, which denotes the ith value of a weight vector w and $f_i(s,a)$, which denotes the value of the ith feature of the Q-state (s,a). He has the following values and observations:

State (s)	Action (a)	New State (s')	Reward (r)
G	Right	D	10
P	Up	Z	1

w_1	w_2	w_3
2	5	10

$f_1(G, Right)$	$f_2(G, Right)$	$f_3(G, Right)$
6	3	4

State	P	Z	D
Q(State, Up)	2	0	0
Q(State, Down)	5	7	0
Q(State, Left)	0	0	2
Q(State, Right)	0	-8	22

(i) [3 pts] What is the initial value of $Q(G, Right)$ based on the above weights and features?

For approximate Q-learning, we compute our desired Q-value by taking a linear combination of the weights and the corresponding features. So we have:

$Q(G, Right) = w_1 \cdot f_1(G, Right) + w_2 \cdot f_2(G, Right) + w_3 \cdot f_3(G, Right)$

$= 2 \cdot 6 + 5 \cdot 3 + 10 \cdot 4$

$= 67$

(ii) [2 pts] What is the resulting weight vector after performing the first iteration of the weight update rule for going right on G? This time, assume a discount factor $\gamma = 0.5$ and learning rate $\alpha = 0.5$.

Recall that the weight update formula for the ith weight is as follows:

$w_i \leftarrow w_i + \alpha \cdot$ difference $\cdot f_i(s,a)$

To perform the weight update, we need to calculate the difference term. The difference term is given as:
difference $= [R(s,a,s') + \gamma \cdot max_{a'} Q(s',a')] - Q(s,a)$
where $Q(s,a)$ is the Q-value we computed using the linear combination of weights and features and $max_{a'} Q(s',a')$ denotes the maximum Q-value for the new state s' by taking a specific action a' on s'.

We compute the difference term using our observation of going right from G to D and our initial $Q(G, Right)$ from the previous part. Plugging in relevant values, we get:
difference $= [10 + 0.5 \cdot 22] - 67 = -46$

Applying this to our weight and feature vectors, we get:
$w_1 = 2 + 0.5 \cdot -46 \cdot 6 = -136$
$w_2 = 5 + 0.5 \cdot -46 \cdot 3 = 5 - 69 = -64$
$w_3 = 10 + 0.5 \cdot -46 \cdot 4 = 10 - 92 = -82$

Our final answer is $[w_1, w_2, w_3] = [-136, -64, -82]$

Q7. [13 pts] Machine Learning: Hotdog vs. Not Hotdog

Bob is building a model to classify whether a picture contains a Hotdog or not. He uses two binary features: whether the picture has brown color in it and whether there is red color in it. He collects this training set:

Brown Color (W_1)	**Red Color** (W_2)	**Label** (y)
1	0	not hotdog
1	0	not hotdog
1	0	not hotdog
1	1	not hotdog
1	1	hotdog
0	0	hotdog

(a) [5 pts] He first builds a Naive Bayes model. Calculate the following probabilities.

(i) [1 pt] $P(y = \text{hotdog}) = \boxed{\frac{1}{3}}$ $\quad P(y = \text{not hotdog}) = \boxed{\frac{2}{3}}$

(ii) [4 pts] $P(W_1 = 1 \mid y = \text{hotdog}) = \boxed{\frac{1}{2}}$ $\quad P(W_1 = 0 \mid y = \text{hotdog}) = \boxed{\frac{1}{2}}$

$P(W_2 = 1 \mid y = \text{hotdog}) = \boxed{\frac{1}{2}}$ $\quad P(W_2 = 0 \mid y = \text{hotdog}) = \boxed{\frac{1}{2}}$

$P(W_1 = 1 \mid y = \text{not hotdog}) = \boxed{1}$ $\quad P(W_1 = 0 \mid y = \text{not hotdog}) = \boxed{0}$

$P(W_2 = 1 \mid y = \text{not hotdog}) = \boxed{\frac{1}{4}}$ $\quad P(W_2 = 0 \mid y = \text{not hotdog}) = \boxed{\frac{3}{4}}$

(b) [3 pts] Next, he uses the model to classify three pictures that are from the test set. Fill in the predicted labels in the table.

Test set

Brown Color (W_1)	**Red Color** (W_2)	**Predicted Label** ($\hat{y}$)
1	1	not hotdog
0	1	hotdog
0	0	hotdog

(c) [5 pts] Bob then adds two new examples to his training set, as shown below.

Additional training examples

Brown Color (W_1)	**Red Color** (W_2)	**Label (y)**
0	0	not hotdog
0	1	not hotdog

(i) [3 pts] Now re-classify the test set using the new larger training set (hint: don't forget to update the prior for each class with the new dataset).

Test set

Brown Color (W_1)	**Red Color (W_2)**	**Predicted Label ($\hat{y}$)**
1	1	not hotdog
0	1	not hotdog
0	0	not hotdog

(ii) [2 pts] Did any of the predictions change? Why?

Yes, the last two predictions switched from hotdog to not hotdog. The main reason is because when the label was not hotdog there were no 0's for W_1 in the training set. So, those two test examples originally had 0 probability for the not hotdog prediction and now have much higher probability.

Q8. [10 pts] Neural Networks: Quadratic Activation Function

Consider the following neural network, where the input is $x \in \mathbb{R}^d$ and the output is a real number prediction $\hat{y} \in \mathbb{R}$. We have two neural network weight matrices: $W_1 \in \mathbb{R}^{n_1 \times n_2}$ and $W_2 \in \mathbb{R}^k$, where k, d are known numbers and you need to find out the value of n_1 and n_2. We consider the following neural network:

$$z = W_1 x; \quad \hat{y} := W_2^T g(z) \tag{29}$$

with a quadratic activation function:

$$g(z) := z \odot z \tag{30}$$

where $\odot$ means element-wise multiplication. For example, $g(\begin{bmatrix}1\\2\\3\end{bmatrix}) = \begin{bmatrix}1\\4\\9\end{bmatrix}$.

We consider a quadratic loss function:

$$\mathcal{L}(W_1; W_2) := \frac{1}{2}(y - \hat{y})^2 \tag{31}$$

(a) [2 pts] What's the value of n_1 and n_2 if the above neural network is valid? You can write the answer in terms of k and d

(i) [1 pt] $n_1 = \boxed{k}$

(ii) [1 pt] $n_2 = \boxed{d}$

(b) [4 pts] Calculate the derivative of the loss function with respect to the following quantities. Your answer should **NOT** include the activation function g (i.e., you need to explicitly calculate the derivative of g rather than writing g'.) On the other hand, your answer can include x, y, $\hat{y}$, W_1, W_2, and z.

(i) [2 pts] $\frac{\partial \mathcal{L}}{\partial W_2} = \boxed{(y - \hat{y})(z) \odot (z)}$

(ii) [2 pts] $\frac{\partial \mathcal{L}}{\partial x} = \boxed{(y - \hat{y}) W_1^T (W_2 \odot 2z)}$

(c) [4 pts] We have realized that the model is not getting the accuracy that we were hoping for. For each of the following possible solutions below, answer yes or no if they could possibly improve the model accuracy.

(i) [1 pt] If the model is overfitting, we can try to make it more complex by increasing the number of layers. $\boxed{\text{No}}$

(ii) [1 pt] If the model is overfitting, we can try to make it simpler by decreasing the number of layers. $\boxed{\text{Yes}}$

(iii) [1 pt] We could try to improve accuracy by getting more training data. $\boxed{\text{Yes}}$

(iv) [1 pt] We could try out different types of models instead of neural networks (e.g., Naive bayes) and see which one works best on the validation set. $\boxed{\text{Yes}}$