

1 Short Questions

~~X~~ When you run Waltz algorithm on the following drawing, which of the following statements is true? Circle the correct answer.

- (i) The algorithm will label all edges uniquely.
- (ii) The algorithm will report that some edges are ambiguous.
- (iii) The algorithm will report that the image cannot be labeled consistently.

ANSWER: (ii)

~~X~~ How many degrees of freedom does a rigid 3-d object have if it moves in a 3-d space?

ANSWER: 6

→ ✓ (c) How does randomized hill-climbing choose the next move each time? Circle the correct answer.

- (i) It generates a random move from the moveset, and accepts this move.
- (ii) It generates a random move from the whole state space, and accepts this move.
- (iii) It generates a random move from the moveset, and accepts this move only if this move improves the evaluation function.
- (iv) It generates a random move from the whole state space, and accepts this move only if this move improves the evaluation function.

ANSWER: (iii)

→ ✓ (d) Suppose you are using a genetic algorithm. Show the children of the following two strings if single point crossover is performed with a cross-point between the 4th and the 5th digits:

1 4 6 2 5 7 2 3 and 8 5 3 4 6 7 6 1

ANSWER: 1 4 6 2 6 7 6 1 and 8 5 3 4 5 7 2 3

✓ (e) What is the entropy of these examples: 1 3 2 3 1 3 3 2

ANSWER: 1.5

→ ✓ (f) Which of the following is the main reason of pruning a decision tree? Circle the correct answer.

- (i) to save computational cost
- (ii) to avoid over-fitting
- (iii) to make the training error smaller

ANSWER: (ii)

→ ✓ (g) Which of the following does the Naive Bayes classifier assume? Circle the correct answer.

- (i) All the attributes are independent.
- (ii) All the attributes are conditionally independent given the output label.
- (iii) All the attributes are jointly dependent to each other.

ANSWER: (ii)

→ ✓ (h) By which of the following networks can XOR function be learned? Circle the correct answer.

- (i) linear perceptron
- (ii) single layer Neural Network
- (iii) 1-hidden layer Neural Network
- (iv) none of the above

ANSWER: (iii)

✗ (i) If we use K-means on a finite set of samples, which of the following statement is true? Circle the correct answer.

- (i) K-means is not guaranteed to terminate.
- (ii) K-means is guaranteed to terminate, but is not guaranteed to find the optimal clustering.
- (iii) K-means is guaranteed to terminate and find the optimal clustering.

ANSWER: (ii)

→ ✓ (j) In the worst case, what is the number of nodes that will be visited by Breadth-First Search in a (non-looping) tree with depth d and branching factor b ?

ANSWER: $O(b^d)$

✗ (k) True or False : If a search tree has cycles, A* Search with an inadmissible heuristic might never converge when run on that tree.

ANSWER: False

→ ✓ (l) Circle the Nash Equilibria in the following matrix-form game:

ANSWER:

		Player 2		
		D	E	F
Player 1	A	0, 1	(3, 5)	2, 1
	B	(6, 3)	1, 3	5, 2
	C	4, 2	3, 4	(7, 7)

→ ✓ (m) Assume the following zero-sum game, where player 1 is the maximizer:

ANSWER:

		Player 2	
		C	D
Player 1	A	2	0
	B	0	1

If Player 1 chooses strategy A with probability p , and if Player 2 always plays strategy C, what is the expected value of the game?

ANSWER: $2p + 0(1 - p) = 2p$

→ ✓ (n) In the mixed strategy Nash equilibrium for the above game, with what probability does Player 1 use strategy A?

$$2p = 1(1 - p)$$

$$3p = 1$$

ANSWER: $p = 1/3$

✗ (o) **True or False** : In a second-price, sealed bid auction, it is optimal to bid your true value. There is no advantage to bluffing.

ANSWER: True

→ ✓ (p) How many values does it take to represent the joint distribution of 4 boolean variables?

ANSWER: 16

→ ✓ (q) If $P(A) = 0.3$, $P(B) = 0.4$, and $P(A|B) = 0.6$

(a) What is $P(A \cap B)$?

$$\text{ANSWER: } P(A \cap B) = P(A|B) * P(B) = 0.24$$

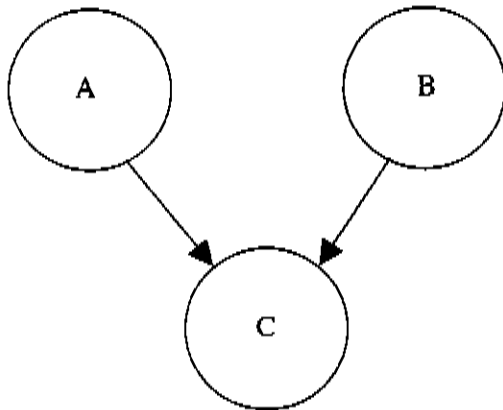
(b) What is $P(B|A)$?

$$\text{ANSWER: } P(B|A) = \frac{P(A|B)P(B)}{P(A)} = 0.8$$

(c) Are A and B independent?

ANSWER: No

X(r) For the following questions, use the diagram below. If you do not have enough information to answer a question, answer False.



- (a) **True or False** : $A \perp B$
ANSWER: True
- (b) **True or False** : $A \perp C$
ANSWER: False
- (c) **True or False** : $I \langle A, \{C\}, B \rangle$
ANSWER: False
- (d) **True or False** : $I \langle C, \{A\}, B \rangle$
ANSWER: False

✓(s) **True or False** : Policy iteration will usually converge to a better policy than value iteration.
ANSWER: False

X(t) **True or False** : For a densely connected MDP with many actions, policy iteration will generally converge faster than value iteration.
ANSWER: True

15-381 Spring 05 Midterm

Tuesday March 1, 2005

Name: _____

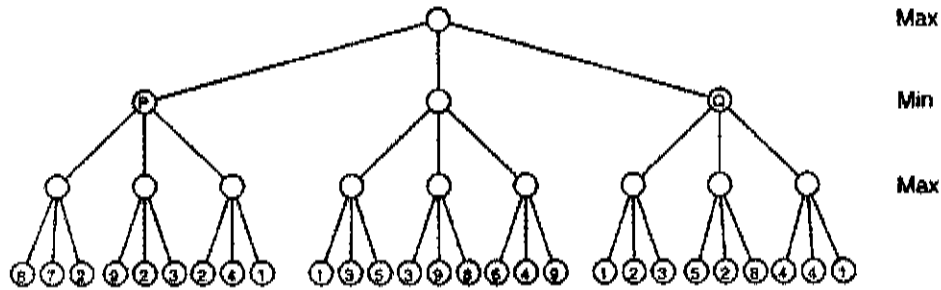
Andrew ID: _____

- This is an open-book, open-notes examination. You have 80 minutes to complete this examination.
- This examination consists of 6 questions, each worth 20 points. For each student, **only the top 5 scoring questions** will be considered. The worst-scoring question will be discarded (thus you may choose to ignore a question and still potentially get full marks). The maximum possible score is 100.
- Write your answers legibly *in the space provided* on the examination sheet. If you use the back of a sheet, indicate clearly that you have done so on the front.
- Write your name and Andrew ID on this page and your andrew id on the top of each successive page in the space provided.
- Calculators are allowed but laptops and PDAs are not allowed.
- Good luck!

	<u>YOU CAN DO</u>
1	✓
2	(a) + (b)
3	✓
4	✓
5	X
6	✓

✓ **1 Problem 1: Game Tree Search (20 pts)**

The figure below is the game tree of a two-player game; the first player is the maximizer and the second player is the minimizer. Use the tree to answer the following questions:



→ (a) What is the final value of this game?

Consider running the alpha-beta pruning algorithm on this game tree.

→ (b) Is the final value of beta at the root node (after all children have been visited) $+\infty$? (T/F)

✓ (c) What is the final value of beta at the node labeled P (after all of P's children have been visited)?

Suppose we are in the middle of running the algorithm. The algorithm has just reached the node labeled Q. The value of alpha is 5 and the value of beta is $+\infty$.

✓ (d) Will any nodes be pruned?

✓ (e) What value will Q return to its parent?

2 Problem 2: Game Theory (20 pts)

Two players, A and B, play a game, in which they each shout out an integer: 1, 2 or 3. If they both shout the same number, they receive a prize:

- They each get one dollar if they both shouted "1".
- They each get two dollars if they both shouted "2".
- They each get three dollars if they both shouted "3".

If they shouted different numbers, they get nothing.

→ (a) Is it a Nash Equilibrium to both shout "1"? (T/F)

Consider the mixed strategy of

- I1: shout "1" with probability $\frac{1}{3}$
 I2: shout "2" with probability $\frac{1}{3}$
 I3: shout "3" with probability $\frac{1}{3}$

→ (b) If both players use this mixed strategy, is that a Nash Equilibrium? (T/F)

Now consider a very different game. Two companies, A and B, both make elbow warmers. The more they spend on advertising, the more sales they get, but there are diminishing returns. A's advertising somewhat helps B, and B's advertising somewhat helps A. The exact formulas are

$$P_a = \overbrace{\log(2a + b)}^{\text{revenue}} - \underbrace{a}_{\text{expense}}$$

$$P_b = \log(a + 2b) - b$$

where

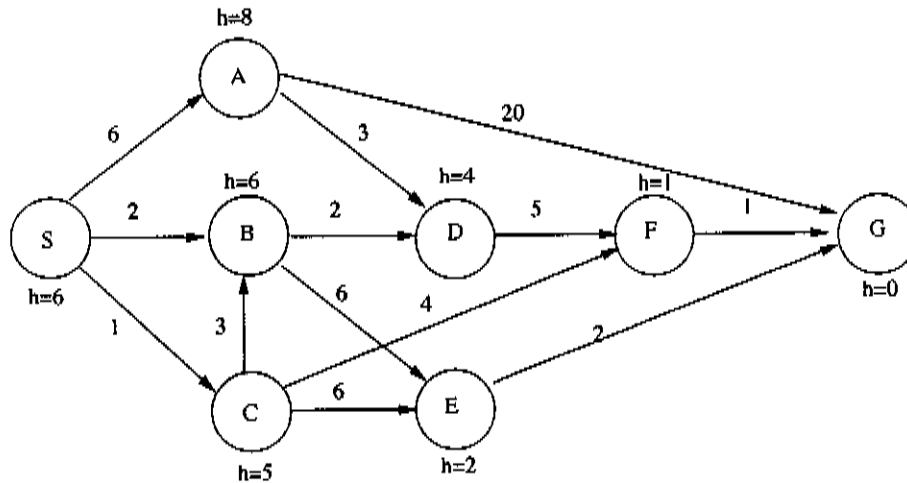
- a = # dollars that A spends on advertising
 b = # dollars that B spends on advertising
 P_a = profit to A
 P_b = profit to B

(c) Work out the Nash equilibrium.

Hint: $\frac{\partial P_a}{\partial a} = \frac{2}{2a+b} - 1$ $\frac{\partial P_a}{\partial b} = \frac{1}{2a+b}$
 $\frac{\partial P_b}{\partial a} = \frac{1}{a+2b}$ $\frac{\partial P_b}{\partial b} = \frac{2}{a+2b} - 1$

3 Problem 3: Search (20 pts)

Consider the search problem below with start state S and goal state G . The transition costs are next to the edges, and the heuristic values are next to the states.



If we use Uniform-Cost Search:

→ (a) What is the final path for this search?

If we use Depth First Search, and it terminates as soon as it reaches the goal state:

√ (b) What is the final path for this DFS search? If a node has multiple successors, then we always expand the successors in increasing alphabetical order.

If we use A* search:

√ (c) What is the final path for this A* search?

→ (d) Is the heuristic function in this example admissible?

4 Problem 4: Hill Climbing, Simulated Annealing and Genetic Algorithm (20 pts)

An N-Queens problem is to place N Queens on an NxN chess board such that no queen attacks any other (a queen can attack any other piece in the same row, column or diagonal). Let's consider one slightly efficient complete-state formulation as below:

- State: All N queens are on the board, one queen per row and per column. In this way, we only need to worry about the attacks along the diagonal, and this simplifies the evaluation function calculation.
- Evaluation: Number of *nonattacking* pairs of queens in this state.
- Successor Function: Swap of *adjacent* columns. For example, swap (1,2) means swap the column#1 and column#2

Let's study the 5-Queens problem:

→ (a) Given the definition above, how many states are there in total?

Number of states:

```

. . Q . .
Q . . . .
. Q . . .
. . . Q .
. . . . Q

```

Figure 1: Initial state

```

. . Q . .   . Q . . .   . . . Q .   . . Q . .
. Q . . .   Q . . . .   Q . . . .   Q . . . .
Q . . . .   . . Q . .   . Q . . .   . Q . . .
. . . Q .   . . . Q .   . . Q . .   . . . Q .
. . . . Q   . . . . Q   . . . . Q   . . . . Q

```

Figure 2: Successors

→ (b) If we carry out steepest ascent hill-climbing starting from the initial state in Figure 1, what is the final state, and is it a solution? (The evaluation function values for the initial state and its four successors are given as below.)

```

InitState: Eval = 8
swap(1,2): Eval = 4
swap(2,3): Eval = 6
swap(3,4): Eval = 6
swap(4,5): Eval = 6

```

√ (c) Consider the relations between simulated annealing and variants of hill-climbing in a general setting:

√ When $T = \infty$, simulated annealing is:

- A. steepest ascent
- B. stochastic hill climbing

- C. first-choice hill climbing
- D. random-restart hill climbing
- E. none of the above

✓When the temperature decay rate = 1, simulated annealing is

- A. steepest ascent
- B. stochastic hill climbing
- C. first-choice hill climbing
- D. random-restart hill climbing
- E. none of the above

6 Problem 6: Constraint Satisfaction (20 pts)

Consider the perennial problem of scheduling classrooms in Wean Hall. We have 4 instructors ($I1, I2, I3, I4$) and 3 rooms ($R1, R2, R3$). We need to assign rooms to instructors. We assume that the instructors need the rooms at the following times:

I1: 9am to 11am

I2: 10am to 2pm

I3: 1pm to 5pm

I4: 1pm to 3pm

We assume that a room can be used by only one instructor at a time and that room $R3$ is too small for instructor $I1$ and that rooms $R2$ and $R3$ are too small for instructor $I3$.

- (a) Show the search with forward checking by writing the domain for each variable at every step in the table below. Write the variable instantiated at each step of the search in the left column and the corresponding value domain for each the variables in the remaining entries of the table. Use the variable ordering ($I1, I2, I3, I4$) and the value ordering ($R1, R2, R3$).

Variable Instantiated	I1	I2	I3	I4
<i>Initial Domains</i>	R1,R2	R1,R2 R3	R1	R1,R2 R3

- √ (b) Can the problem be solved by constraint satisfaction alone without backtracking?
propagation

15-381 Spring 05 Final

Thursday May 5th, 2005

Name: _____ Andrew ID: _____

- This is an open-book, open-notes examination. You have 180 minutes (3 hours) to complete this examination.
- Write your answers legibly *in the space provided* on the examination sheet. If you use the back of a sheet, indicate clearly that you have done so on the front.
- Write your name and Andrew ID on this page and your Andrew ID on the top of each successive page in the space provided.
- There are 10 questions in this exam. Some of them are marked with the words "more difficult". Please budget your time using this information.
- Calculators are allowed, but laptops and PDAs are not allowed.
- Good luck!

YOU CAN DO

Question	Score	
✓ 1. Game Tree Search	10	
X 2. Uninformed Search	10	
✓ 3. Decision Tree	14	
✓ 4. Naive Bayes	10	
X 5. K-Means	7	
X 6. Cross Validation	10	
✓ 7. Bayes Net	10	
✓ 8. Markov Decision Process	9	
(a) + (b) ✓ 9. Neural Net	12	
✓ 10. Reinforcement Learning	8	
Total	100	

1 Game Tree Search (10 points)

Figure 1 shows the game tree of a two-player game; the first player is the maximizer and the second player is the minimizer. Use the tree to answer the following questions.

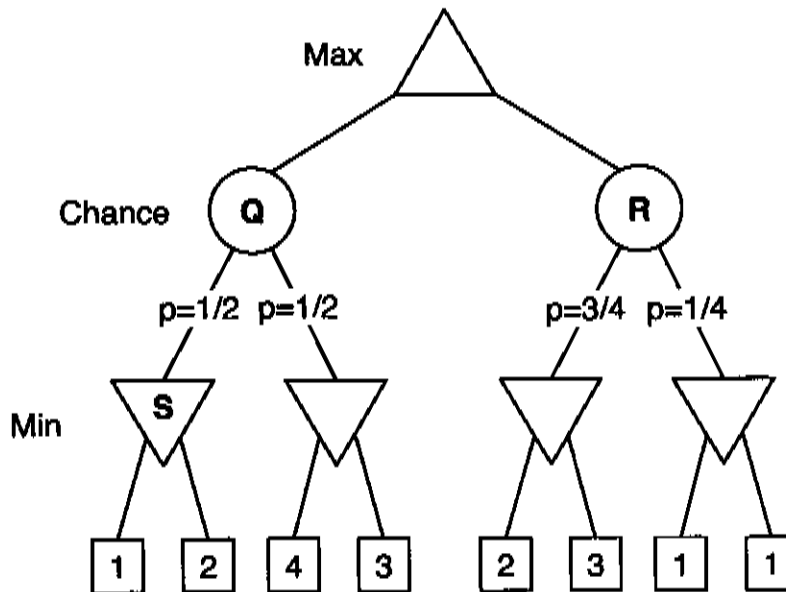


Figure 1: Game tree of two-player game with chance

→ 1. Circle one (2 pts) : What is the value of the node labeled S?
Answer: (b)

- (a) $1/2$
- (b) 1
- (c) 2
- (d) Not enough information / cannot be determined

→ 2. Circle one (2 pts): What is the expected value of the node labeled Q?
Answer: (d)

- (a) $1/2$
- (b) 1
- (c) $3/2$
- (d) 2
- (e) 3
- (f) Not enough information / cannot be determined

✓ 3. **Circle one (2 pts):** What is the expected value of the node labeled R?

Answer: (b)

(a) 2

(b) $7/4$

(c) 1

(d) $2/3$

(e) $1/4$

(f) Not enough information / cannot be determined

✓ 4. **Circle one (2 pts):** What is the expected value of the game?

Answer: (c)

(a) 1

(b) $7/4$

(c) 2

(d) 3

(e) Not enough information / cannot be determined

✗ 5. **True or False (2 pts):** You have been provided with enough information so that you could modify the alpha-beta pruning algorithm to work on this game tree.

Answer: True

3 Decision Trees (14 points)

Your spaceship has just landed on an alien planet, and your crew has begun investigating the local wildlife. Unfortunately, most of your scientific equipment is broken, so all you can tell about a given object is what color it is, how many eyes it has, and whether or not it is alive. To make matters worse, none of you are biologists, so you are going to have to use a decision tree to classify objects near your landing site as either alive or not alive. Use the table below to answer the following questions:

Object	Color	Number of eyes	Alive
A	Red	4	Yes
B	Black	42	No
C	Red	13	Yes
D	Green	3	Yes
E	Black	27	No
F	Red	2	Yes
G	Black	1	Yes
H	Green	11	No

1. Circle one (2 pts): Which of the following is the largest? (Note that we are not asking for exact values. You may solve this problem by simply inspecting the table.)

Answer: (c)

- (a) $H(\text{Alive}|\text{Number of eyes} > 10)$
- (b) $H(\text{Alive}|\text{Number of eyes} < 5)$
- (c) $H(\text{Alive}|\text{Color} = \text{Green})$
- (d) $H(\text{Alive}|\text{Color} = \text{Black})$

2. Fill in the following blank (2 pts):

What is the entropy of Alive?

Answer: $H(\text{Alive}) = -\frac{5}{8}\log_2\frac{5}{8} - \frac{3}{8}\log_2\frac{3}{8} = 0.9544$

3. Fill in the following blank (2 pts):

What is $IG(\text{Alive}|\text{Color})$?

Answer: 0.36

4. Circle one (2 pts): Suppose we wanted to turn Number of eyes into a binary attribute for the purpose of building a decision tree. Which of the following binary categorical splits results in the larger value of $IG(\text{Alive}|\text{Number of eyes})$? (Note that we are not asking for exact values. You may solve this problem by simply inspecting the table.)

Answer: (b)

- (a) {Number of eyes = 11, Number of eyes \neq 11}
- (b) {Number of eyes \leq 4, Number of eyes $>$ 4}
- (c) {Number of eyes \leq 13, Number of eyes $>$ 13}

✓ 5. (6 pts) Suppose we were going to build a decision tree for this data:

- First, we split using the attribute you chose in the previous question.
- Second, we split on Color.

How would this tree classify the following objects? (In case of a tie at a leaf node, classify the object as Not alive.) NOTE: This should not be a very complicated tree.

- (a) Circle one (3 pts): **(Alive or Not alive)** A red object with 23 eyes
- (b) Circle one (3 pts): **(Alive or Not alive)** A black object with 1.5 eyes

Answer:

If (a) on previous question, Not Alive; Not Alive

If (b) on previous question, Alive; Alive

If (c) on previous question, Not Alive ; Alive

4 Naive Bayes (10 points)

For each question below, you are **required to write down the basic formulas that you use to compute your answers**. Otherwise, you can only get a maximum of half credit.

Tom is a CMU student. Recently, his mood has been highly influenced by two factors: the weather (W) and his study (S). Naturally, he likes good weather and hates bad weather. More importantly, Tom worries about his exams. Tom feels happy if he passes exams and not happy if he fails them. Now Tom wants to predict his happiness according to these two factors using his previous experience. Tables A and B show this data.

Weather(W)	Study(S)	Happy(H)
Bad	Fail	0
Good	Fail	0
Good	Fail	0
Good	Fail	0
Bad	Pass	0
Bad	Pass	1
Bad	Pass	1
Good	Pass	1

Table A: 2 factors

- (a) Using Table A: If today's situation is W=Good, S=Pass, and Tom uses a Naive Bayes classifier, how would he predict his happiness? Please show your computations and the classifier's prediction. (2 pts)

Answer:

$$P(W = G|H = 0)P(S = P|H = 0)P(H = 0) = 3/40$$

$$P(W = G|H = 1)P(S = P|H = 1)P(H = 1) = 1/8$$

predict Happy.

- √ (b) Using Table A: If today's situation is W=Bad, S=Fail, and Tom uses a Naive Bayes classifier, how would he predict his happiness? Please show your computations and the classifier's prediction. (2 pts)

Answer:

$$P(W = B|H = 0)P(S = F|H = 0)P(H = 0) = 1/5$$

$$P(W = B|H = 1)P(S = F|H = 1)P(H = 1) = 0$$

predict Unhappy.

Tom also notices that his neighbor always goes for a walk if the weather is good and stays at home if the weather is bad. Tom thinks it wouldn't hurt to have more information, so he adds one more

factor, Neighbor (N), to the table. The new table is shown as Table B. You can see that whenever $W=Good$, $N=Out$, and whenever $W=Bad$, $N=home$.

Weather(W)	Study(S)	Neighbor(N)	Happy(H)
Bad	Fail	Home	0
Good	Fail	Out	0
Good	Fail	Out	0
Good	Fail	Out	0
Bad	Pass	Home	0
Bad	Pass	Home	1
Bad	Pass	Home	1
Good	Pass	Out	1

Table B: 3 factors

√ (c) Using Table B: Now, if $W=Good$, $S=Pass$, $N=Out$, and Tom uses a Naive Bayes Classifier, how would he predict his happiness? Please show your computations and the classifier's prediction. (2 pts)

Answer:

$$P(W = G|H = 0)P(S = P|H = 0)P(H = 0) = 9/200$$

$$P(W = G|H = 1)P(S = P|H = 1)P(H = 1) = 1/24$$

predict Unhappy.

HAZARD → (d) Will the new factor improve the performance of the Naive Bayes classifier? Why or why not? (2 pts)

Answer: No. Since weather and neighbor are not conditionally independent. The assumption of Naive Bayes doesn't hold anymore.

→ (e) Using Table B: Now, if Tom uses a Bayes Classifier instead of a Naive Bayes Classifier, and we still assume $W=Good$, $S=Pass$, $N=Out$, how would he predict his happiness? Please show your computations and the classifier's prediction. (2 pts)

Answer:

$$P(W = G, S = P, N = O|H = 0)P(H = 0) = 0$$

$$P(W = G, S = P, N = O|H = 1)P(H = 1) = 1/8$$

predict Happy.

7 Bayes Nets (10 points)

Given the Bayes net shown in the Figure below; A, B, C, D, and E are all Boolean variables. $P(A=“+”)$ is simply denoted as $P(A)$.

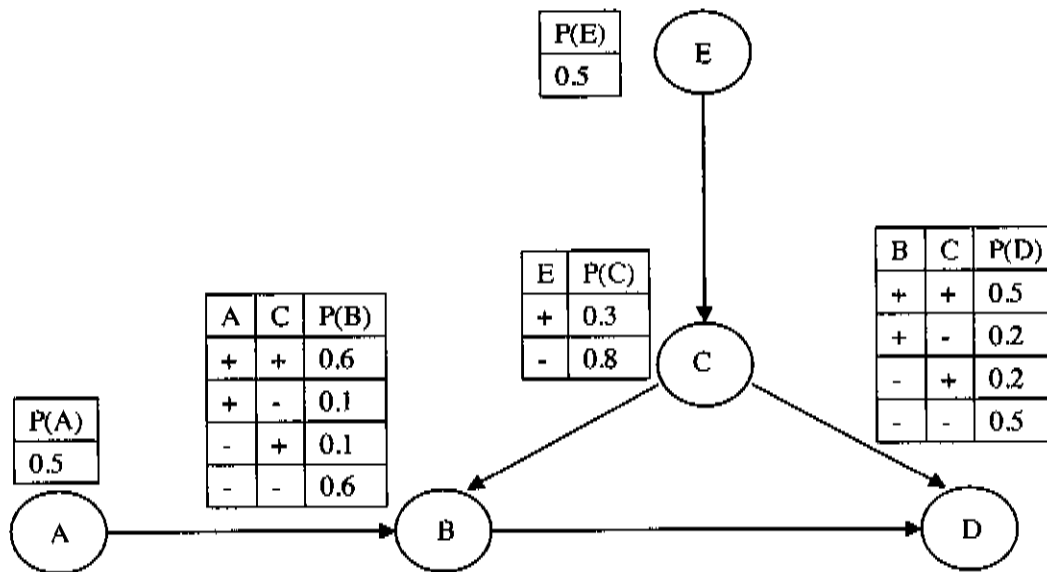


Figure 2: Bayes Net

Note: In the following, we'll use the notations: $A \perp B$ means A is independent of B; $A \perp B|C$ means A is conditionally independent of B given C.

→ (a) Please judge if the following independence assumptions are correct or not:

1. (True or False) (1 pt): $B \perp E|C$

Answer: True

2. (True or False) (1 pt): $A \perp D$

Answer: False

3. (True or False) (2 pts): $A \perp D|B$

Answer: False

4. (True or False) (2 pts): $A \perp D|B, C$

Answer: True

→ (b) Compute the value of $P(C)$ (2 pts)

Answer: 0.55

→ (c) Compute the value of $P(B|A)$ (2 pts)

Answer: 0.375

8 Markov Decision Process (9 points)

In the following Markov Decision Process, there are three states S_1 , S_2 and S_3 . The rewards for each state and all of the state transitions are marked in the given figure. There are two actions. Action a_1 causes you to stay in the same state, and action a_2 causes you to move to other states. The discount factor is $\gamma = 0.9$.

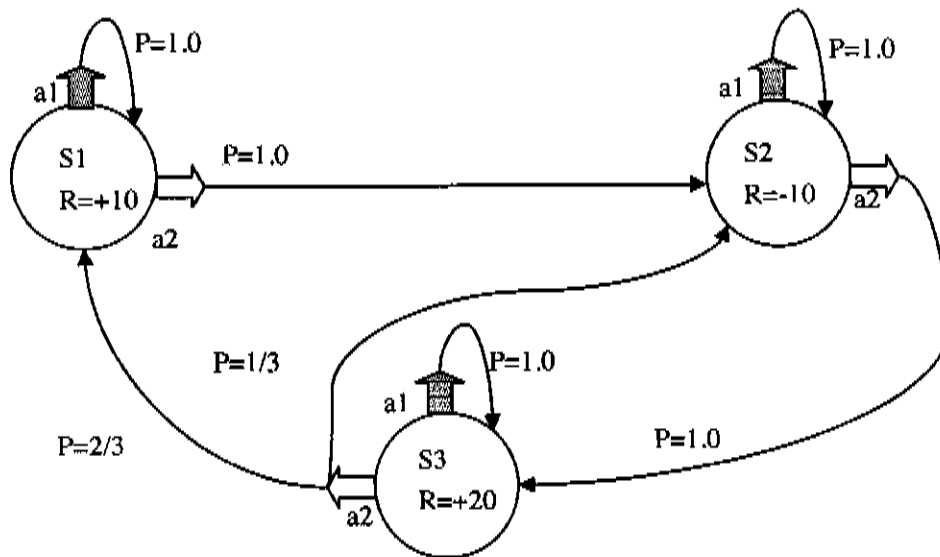


Figure 3: Markov Decision Process

→ (a) How many distinct policies are there in the above MDP? (1 pt)
 Answer: $2^3 = 8$

√ (b) $U^{\pi_0}(i)$ is the expected sum of discounted rewards if we start at state i and follow the policy π_0 . The initial policy π_0 is the one that assigns a_1 to every state. Write down the numerical values of the expected discounted rewards of each of the following states in the MDP. (3 pts)

→ 1. $U^{\pi_0}(S_1) = 100$

2. $U^{\pi_0}(S_2) = -100$

3. $U^{\pi_0}(S_3) = 200$

→ (c) Continuing from part (b), suppose we run policy iteration with π_0 as the initial policy. Define π_1 as the updated policy after one iteration of policy iteration, and write down the updated policy for each of the states (2 pts):

1. $\pi_1(S_1) = a_1$

2. $\pi_1(S_2) = a_2$

3. $\pi_1(S_3) = a_1$

√ (d) Suppose we run value iteration, $U^k(i)$ is the expected sum of discounted rewards if we start at state i after $(k - 1)$ -step of value iterations. Starting from the initial value of $U^1(i)$, which is the reward at state i , please write down the updated $U^2(i)$ after one value iteration (3 pts).

- 1. $U^2(S_1) = 19$
2. $U^2(S_2) = 8$
3. $U^2(S_3) = 38$

9 Neural Network (12 points)

Consider a problem in which each data point has two coordinates u and v . We wish to learn a classifier for this problem by using a linear perceptron network with inputs u and v and weights W_u and W_v on the two connections. We use a threshold of 0 so that the output of the network is +1 (class 1) if the output unit is greater than or equal to 0 and -1 (class 2) otherwise.

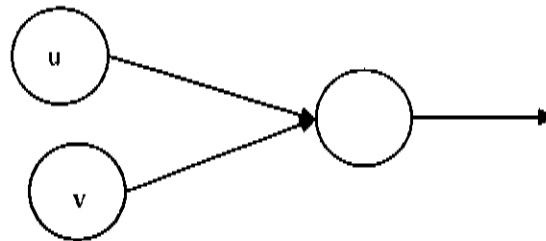
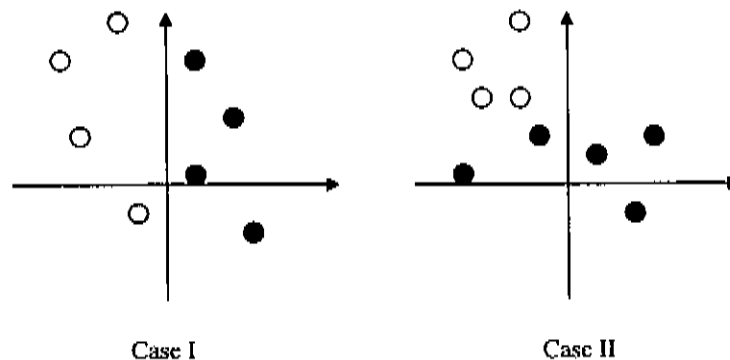


Figure 4: Neural Net

- (a) Can the network distinguish the two classes in the cases illustrated below? Why/Why not in each case? (2 pts)



Answer:

Yes for case 1 and no for case 2. Without a constant term, the decision boundary must go through the origin.

- (b) Assume that we augment the network with an additional input unit with constant value 1 (we'll call the weight on the additional connection W). How does the answer to 1 change and why? (2 pts)

Answer:

With a constant term the decision boundary can go anywhere though it must still be linear. So both cases I and II can now be discriminated correctly.

~~(c) (More difficult) The sigmoid perceptron defines Prediction = $g(w_0 + w_1x_1 + w_2x_2 + \dots + w_kx_k)$, given input variables x_1 through x_k , and $g(z) = \frac{1}{1 + \exp(-z)}$. Pat thinks the sigmoid function is too computationally expensive and replaces it with this piecewise linear approximation: Prediction =~~

√ 10 Reinforcement learning (8 points)

Part I.

Consider an agent starting in a room A in which it can take two possible actions: to leave the room (action " L ") or to stay (action " S "). If it leaves A , the agent moves to room B , which is a terminal state (no more actions can be taken). The outcomes of the actions are uncertain, so that when executing action L (or action S), there is some probability that the agent will leave A (or stay in A). We assume that the reward in entering state B is $R(B) = +1$ and the reward for being in state A is $R(A) = -0.1$.

- √ (a) Draw the (very simple) diagram corresponding to this MDP. Answer by inspection of the diagram: What is the optimal policy? (2 pts)



Answer: With the added conditions on the probabilities, the optimal policy is $\pi(A) = L$.

- √ (b) Assume that the agent knows neither the world (transition probabilities) nor the utilities of the states. Assume that the agent, for some reason, happens to follow the optimal policy. The rewards received at states A and B are the same as described above. In the process of executing this policy, the agent execute four trials and, in each trial, it stops after reaching state B . The following state sequences are recorded during the trials: $AAAB$, AAB , AB , AB . What is the estimate of $T(., ., .)$? What is the estimate of $U(A)$, assuming a discount factor of $\gamma = 0.5$? (2 pts)

Answer:

$$T(A, L, A) = 3/7 \text{ and } T(A, L, B) = 4/7$$

Note that $T(A, S, A)$ cannot be computed from the data given in the text and it is not needed since we assume that we follow the optimal policy.

$$U(A) = R(A) + \gamma(T(A, L, A)U(A) + T(A, L, B)U(B))$$

$$U(A) = -0.1 + 0.5 \times (3/7 \times U(A) + 4/7 \times 1)$$

$$11/14 \times U(A) = -0.1 + 4/14$$

$$U(A) = 26/110 = 0.2364$$

- √ (c) Assume now that the agent is executing only one trial yielding the sequence of states AAB . Compute the estimate of the utility $U(A)$ using TD (temporal differencing). Use discount $\gamma = 0.5$, and learning rate $\alpha = 0.5$. (2 pts)

Answer:

Transition A to A :

$$U^{new}(A) = U^{old}(A) + \alpha(R(A) + \gamma U^{old}(A) - U^{old}(A))$$

$$U^{new}(A) = -0.1 + 0.5 \times (-0.1 + 0.5 \times -0.1 - (-0.1)) = -0.125$$

Transition A to B :

$$U^{new}(A) = U^{old}(A) + \alpha(R(A) + \gamma U(B) - U^{old}(A))$$

$$U^{new}(A) = -0.125 + 0.5 \times (-0.1 + 0.5 \times 1 - (-0.125)) = 0.1375$$

Note that the question did not specify the starting values for U . Alternative solutions (e.g., with $U = 0$) were also accepted as long as the formulas were correct.

Part II.

We are using Q-learning to learn a policy in an MDP with two states S_1 and S_2 and two actions a and b . Assume that $\gamma = 0.8$ and $\alpha = 0.2$, and that the current values of Q are:

$Q(S_1, a)$	2.0
$Q(S_1, b)$	2.0
$Q(S_2, a)$	4.0
$Q(S_2, b)$	2.0

Suppose that, when we were in state S_1 , we took action b , received reward 1.0 and moved to state S_2 . Which item of the Q table will change and what is the new value? (2 pts)

Answer:

$Q(S_1, b)$ is the affected entry.

$$Q^{new}(S_1, b) = Q^{old}(S_1, b) + \alpha(R(S_1) + \gamma \text{Max}_{action} Q(S_2, action) - Q^{old}(S_1, b))$$

$$Q^{new}(S_1, b) = 2 + 0.2 \times (1 + 0.8 \times 4 - 2) = 2.44$$

Note: A common mistake is to forget the "Max" and to use 0.8×2 instead of the correct expression.

$$A_{i,j} = \frac{(i-j)}{2}$$

A	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	.5000	0	0	0	0	0	0	0	0	0	0
3	0	.3333	.6667	0	0	0	0	0	0	0	0	0
4	0	.2500	.5000	.7500	0	0	0	0	0	0	0	0
5	0	.2000	.4000	.6000	.8000	0	0	0	0	0	0	0
6	0	.1667	.3333	.5000	.6667	.8333	0	0	0	0	0	0
7	0	.1429	.2857	.4286	.5714	.7143	.8571	0	0	0	0	0
8	0	.1250	.2500	.3750	.5000	.6250	.7500	.8750	0	0	0	0
9	0	.1111	.2222	.3333	.4444	.5556	.6667	.7778	.8889	0	0	0
10	0	.1000	.2000	.3000	.4000	.5000	.6000	.7000	.8000	.9000	0	0
11	0	.0909	.1818	.2727	.3636	.4545	.5455	.6364	.7273	.8182	.9091	0
12	0	.0833	.1667	.2500	.3333	.4167	.5000	.5833	.6667	.7500	.8333	.9167

$$H_{i,j} = I(A_{i,j}) - I(A_{i,j})$$

H	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1.0000	0	0	0	0	0	0	0	0	0	0
3	0	.9183	.9183	0	0	0	0	0	0	0	0	0
4	0	.8113	1.0000	.8113	0	0	0	0	0	0	0	0
5	0	.7219	.9710	.9710	.7219	0	0	0	0	0	0	0
6	0	.6500	.9183	1.0000	.9183	.6500	0	0	0	0	0	0
7	0	.5917	.8631	.9852	.9852	.8631	.5917	0	0	0	0	0
8	0	.5436	.8113	.9344	1.0000	.9344	.8113	.5436	0	0	0	0
9	0	.5033	.7642	.9183	.9911	.9911	.9183	.7642	.5033	0	0	0
10	0	.4690	.7219	.8813	.9710	1.0000	.9710	.8813	.7219	.4690	0	0
11	0	.4395	.6840	.8454	.9457	.9940	.9940	.9457	.8454	.6840	.4395	0
12	0	.4138	.6500	.8113	.9183	.9799	1.0000	.9799	.9183	.8113	.6500	.4138

$$H_{i,j} = I(A_{i,j}) - I(A_{i,j})$$

$$= I\left(\frac{i-1}{2}, \frac{i-1}{2}\right)$$

Binary Entropy Table

	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.00	0.0000	0.0014	0.0028	0.0042	0.0056	0.0070	0.0084	0.0098	0.0112	0.0126
0.01	0.0098	0.0112	0.0126	0.0140	0.0154	0.0168	0.0182	0.0196	0.0210	0.0224
0.02	0.0154	0.0168	0.0182	0.0196	0.0210	0.0224	0.0238	0.0252	0.0266	0.0280
0.03	0.0182	0.0196	0.0210	0.0224	0.0238	0.0252	0.0266	0.0280	0.0294	0.0308
0.04	0.0210	0.0224	0.0238	0.0252	0.0266	0.0280	0.0294	0.0308	0.0322	0.0336
0.05	0.0238	0.0252	0.0266	0.0280	0.0294	0.0308	0.0322	0.0336	0.0350	0.0364
0.06	0.0266	0.0280	0.0294	0.0308	0.0322	0.0336	0.0350	0.0364	0.0378	0.0392
0.07	0.0294	0.0308	0.0322	0.0336	0.0350	0.0364	0.0378	0.0392	0.0406	0.0420
0.08	0.0322	0.0336	0.0350	0.0364	0.0378	0.0392	0.0406	0.0420	0.0434	0.0448
0.09	0.0350	0.0364	0.0378	0.0392	0.0406	0.0420	0.0434	0.0448	0.0462	0.0476
0.10	0.0378	0.0392	0.0406	0.0420	0.0434	0.0448	0.0462	0.0476	0.0490	0.0504
0.11	0.0406	0.0420	0.0434	0.0448	0.0462	0.0476	0.0490	0.0504	0.0518	0.0532
0.12	0.0434	0.0448	0.0462	0.0476	0.0490	0.0504	0.0518	0.0532	0.0546	0.0560
0.13	0.0462	0.0476	0.0490	0.0504	0.0518	0.0532	0.0546	0.0560	0.0574	0.0588
0.14	0.0490	0.0504	0.0518	0.0532	0.0546	0.0560	0.0574	0.0588	0.0602	0.0616
0.15	0.0518	0.0532	0.0546	0.0560	0.0574	0.0588	0.0602	0.0616	0.0630	0.0644
0.16	0.0546	0.0560	0.0574	0.0588	0.0602	0.0616	0.0630	0.0644	0.0658	0.0672
0.17	0.0574	0.0588	0.0602	0.0616	0.0630	0.0644	0.0658	0.0672	0.0686	0.0700
0.18	0.0602	0.0616	0.0630	0.0644	0.0658	0.0672	0.0686	0.0700	0.0714	0.0728
0.19	0.0630	0.0644	0.0658	0.0672	0.0686	0.0700	0.0714	0.0728	0.0742	0.0756
0.20	0.0658	0.0672	0.0686	0.0700	0.0714	0.0728	0.0742	0.0756	0.0770	0.0784
0.21	0.0686	0.0700	0.0714	0.0728	0.0742	0.0756	0.0770	0.0784	0.0798	0.0812
0.22	0.0714	0.0728	0.0742	0.0756	0.0770	0.0784	0.0798	0.0812	0.0826	0.0840
0.23	0.0742	0.0756	0.0770	0.0784	0.0798	0.0812	0.0826	0.0840	0.0854	0.0868
0.24	0.0770	0.0784	0.0798	0.0812	0.0826	0.0840	0.0854	0.0868	0.0882	0.0896
0.25	0.0798	0.0812	0.0826	0.0840	0.0854	0.0868	0.0882	0.0896	0.0910	0.0924
0.26	0.0826	0.0840	0.0854	0.0868	0.0882	0.0896	0.0910	0.0924	0.0938	0.0952
0.27	0.0854	0.0868	0.0882	0.0896	0.0910	0.0924	0.0938	0.0952	0.0966	0.0980
0.28	0.0882	0.0896	0.0910	0.0924	0.0938	0.0952	0.0966	0.0980	0.0994	0.1008
0.29	0.0910	0.0924	0.0938	0.0952	0.0966	0.0980	0.0994	0.1008	0.1022	0.1036
0.30	0.0938	0.0952	0.0966	0.0980	0.0994	0.1008	0.1022	0.1036	0.1050	0.1064
0.31	0.0966	0.0980	0.0994	0.1008	0.1022	0.1036	0.1050	0.1064	0.1078	0.1092
0.32	0.0994	0.1008	0.1022	0.1036	0.1050	0.1064	0.1078	0.1092	0.1106	0.1120
0.33	0.1022	0.1036	0.1050	0.1064	0.1078	0.1092	0.1106	0.1120	0.1134	0.1148
0.34	0.1050	0.1064	0.1078	0.1092	0.1106	0.1120	0.1134	0.1148	0.1162	0.1176
0.35	0.1078	0.1092	0.1106	0.1120	0.1134	0.1148	0.1162	0.1176	0.1190	0.1204
0.36	0.1106	0.1120	0.1134	0.1148	0.1162	0.1176	0.1190	0.1204	0.1218	0.1232
0.37	0.1134	0.1148	0.1162	0.1176	0.1190	0.1204	0.1218	0.1232	0.1246	0.1260
0.38	0.1162	0.1176	0.1190	0.1204	0.1218	0.1232	0.1246	0.1260	0.1274	0.1288
0.39	0.1190	0.1204	0.1218	0.1232	0.1246	0.1260	0.1274	0.1288	0.1302	0.1316
0.40	0.1218	0.1232	0.1246	0.1260	0.1274	0.1288	0.1302	0.1316	0.1330	0.1344
0.41	0.1246	0.1260	0.1274	0.1288	0.1302	0.1316	0.1330	0.1344	0.1358	0.1372
0.42	0.1274	0.1288	0.1302	0.1316	0.1330	0.1344	0.1358	0.1372	0.1386	0.1400
0.43	0.1302	0.1316	0.1330	0.1344	0.1358	0.1372	0.1386	0.1400	0.1414	0.1428
0.44	0.1330	0.1344	0.1358	0.1372	0.1386	0.1400	0.1414	0.1428	0.1442	0.1456
0.45	0.1358	0.1372	0.1386	0.1400	0.1414	0.1428	0.1442	0.1456	0.1470	0.1484
0.46	0.1386	0.1400	0.1414	0.1428	0.1442	0.1456	0.1470	0.1484	0.1498	0.1512
0.47	0.1414	0.1428	0.1442	0.1456	0.1470	0.1484	0.1498	0.1512	0.1526	0.1540
0.48	0.1442	0.1456	0.1470	0.1484	0.1498	0.1512	0.1526	0.1540	0.1554	0.1568
0.49	0.1470	0.1484	0.1498	0.1512	0.1526	0.1540	0.1554	0.1568	0.1582	0.1596
0.50	0.1498	0.1512	0.1526	0.1540	0.1554	0.1568	0.1582	0.1596	0.1610	0.1624