CS 188 Spring 2012 | Introduction to Artificial Intelligence | Practice Midterm II Solutions

Q1. [19 pts] Cheating at cs188-Blackjack

Cheating dealers have become a serious problem at the cs188-Blackjack tables. A cs188-Blackjack deck has 3 card types (5,10,11) and an honest dealer is equally likely to deal each of the 3 cards. When a player holds 11, cheating dealers deal a 5 with probability $\frac{1}{4}$, 10 with probability $\frac{1}{2}$, and 11 with probability $\frac{1}{4}$. You estimate that $\frac{4}{5}$ of the dealers in your casino are honest (H) while $\frac{1}{5}$ are cheating ($\neg H$).

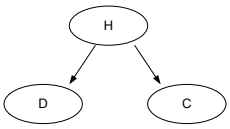
Note: You may write answers in the form of arithmetic expressions involving numbers or references to probabilities that have been directly specified, are specified in your conditional probability tables below, or are specified as answers to previous questions.

(a) [3 pts] You see a dealer deal an 11 to a player holding 11. What is the probability that the dealer is cheating?

$$P(\neg H|D=11) = \frac{P(\neg H, D=11)}{P(D=11)} = \frac{P(D=11|\neg H)P(\neg H)}{P(D=11|\neg H)P(\neg H) + P(D=11|H)P(H)} = \frac{(1/4)(2/10)}{(1/4)(2/10)+(1/3)(8/10)} = 3/19$$

The casino has decided to install a camera to observe its dealers. Cheating dealers are observed doing suspicious things on camera (C) $\frac{4}{5}$ of the time, while honest dealers are observed doing suspicious things $\frac{1}{4}$ of the time.

(b) [4 pts] Draw a Bayes net with the variables H (honest dealer), D (card dealt to a player holding 11), and C (suspicious behavior on camera). Write the conditional probability tables.



H	$P(H)$
1	0.8
0	0.2

H	D	$P(D\|H)$
1	5	1/3
1	10	1/3
1	11	1/3
0	5	1/4
0	10	1/2
0	11	1/4

H	C	$P(C\|H)$
1	1	1/4
1	0	3/4
0	1	4/5
0	0	1/5

(c) [2 pts] List all conditional independence assertions made by your Bayes net.

$$D \perp\!\!\!\perp C|H$$

Common mistakes:

- Stating that two variables are NOT independent; Bayes nets do not guarantee that variables are dependent. This can only be verified by examining the exact probability distributions.

(d) [4 pts] What is the probability that a dealer is honest given that he deals a 10 to a player holding 11 and is observed doing something suspicious?

$$\begin{aligned} P(H|D=10,C) &= \frac{P(H,D=10,C)}{P(D=10,C)} \\ &= \frac{P(H)P(D=10|H)P(C|H)}{P(H)P(D=10|H)P(C|H)+P(\neg H)P(D=10|\neg H)P(C|\neg H)} \\ &= \frac{(4/5)(1/3)(1/4)}{(4/5)(1/3)(1/4)+(1/5)(1/2)(4/5)} \\ &= \frac{5}{11} \end{aligned}$$

Common mistakes:

- -1 for not giving the proper form of Bayes rule, $P(H|D=10,C) = P(H,D=10,C)/P(D=10,C)$
- -1 For a correctly drawn Bayes net C and D are not independent, which means that $P(D=10,C) \neq P(D=10)P(C)$

You can either arrest dealers or let them continue working. If you arrest a dealer and he turns out to be cheating, you will earn a \$4 bonus. However, if you arrest the dealer and he turns out to be innocent, he will sue you for -\$10. Allowing the cheater to continue working will cost you -\$2, while allowing an honest dealer to continue working will get you \$1. Assume a linear utility function $U(x) = x$.

(e) [3 pts] You observe a dealer doing something suspicious (C) and also observe that he deals a 10 to a player holding 11. Should you arrest the dealer?

Arresting the dealer yields an expected payoff of

$$4 * P(\neg H|D=10,C) + (-10) * P(H|D=10,C) = 4(6/11) + (-10)(5/11) = -26/11$$

Letting him continue working yields a payoff of

$$(-2) * P(\neg H|D=10,C) + 1 * P(H|D=10,C) = (-2)(6/11) + (1)(5/11) = -7/11$$

Therefore, you should let the dealer continue working.

(f) [3 pts] A private investigator approaches you and offers to investigate the dealer from the previous part. If you hire him, he will tell you with 100% certainty whether the dealer is cheating or honest, and you can then make a decision about whether to arrest him or not. How much would you be willing to pay for this information?

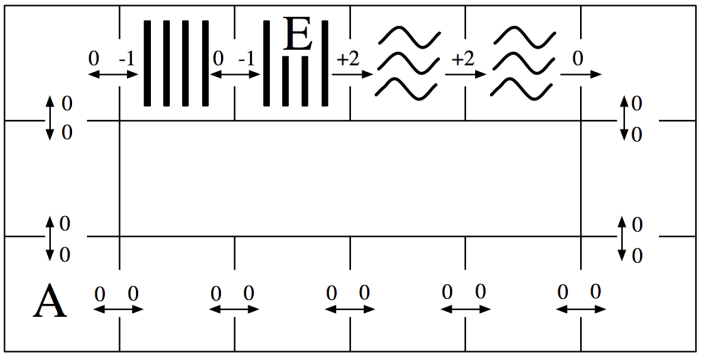
If you hire the private investigator, if the dealer is a cheater you can arrest him for a payoff of \$4. If he is an honest dealer you can let him continue working for a payoff of \$1. The benefit from hiring the investigator is therefore

$$(4) * P(\neg H|D = 10, C) + 1 * P(H|D = 10, C) = 4(6/11) + (1)(5/11) = 29/11$$

If you do not hire him, your best course of action is to let the dealer continue working for an expected payoff of $-7/11$. Therefore, you are willing to pay up to $29/11 - (-7/11) = 36/11$ to hire the investigator.

Q2. [20 pts] MDPs: Grid-World Water Park

Consider the MDP drawn below. The state space consists of all squares in a grid-world water park. There is a single waterslide that is composed of two ladder squares and two slide squares (marked with vertical bars and squiggly lines respectively). An agent in this water park can move from any square to any neighboring square, unless the current square is a slide in which case it must move forward one square along the slide. The actions are denoted by arrows between squares on the map and all deterministically move the agent in the given direction. The agent cannot stand still: it must move on each time step. Rewards are also shown below: the agent feels great pleasure as it slides down the water slide (+2), a certain amount of discomfort as it climbs the rungs of the ladder (-1), and receives rewards of 0 otherwise. The time horizon is infinite; this MDP goes on forever.



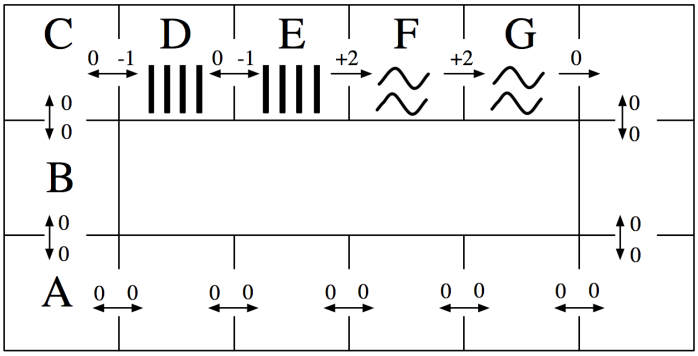
(a) [2 pts] How many (deterministic) policies π are possible for this MDP?

2^{11}

(b) [9 pts] Fill in the blank cells of this table with values that are correct for the corresponding function, discount, and state. Hint: You should not need to do substantial calculation here.

	γ	$s = A$	$s = E$
$V_3^*(s)$	1.0	0	4
$V_{10}^*(s)$	1.0	2	4
$V_{10}^*(s)$	0.1	0	2.2
$Q_1^*(s, \text{left})$	1.0	———	0
$Q_{10}^*(s, \text{left})$	1.0	———	3
$V^*(s)$	1.0	∞	∞
$V^*(s)$	0.1	0	2.2

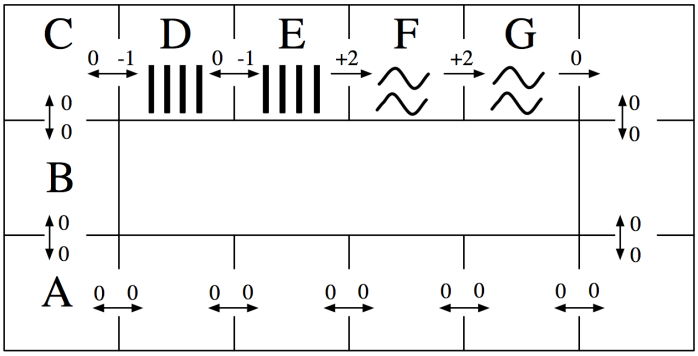
Use this labeling of the state space to complete the remaining subproblems:



(c) [5 pts] Fill in the blank cells of this table with the Q-values that result from applying the Q-update for the transition specified on each row. You may leave Q-values that are unaffected by the current update blank. Use discount $\gamma = 1.0$ and learning rate $\alpha = 0.5$. Assume all Q-values are initialized to 0. (Note: the specified transitions would not arise from a single episode.)

	$Q(D, \text{left})$	$Q(D, \text{right})$	$Q(E, \text{left})$	$Q(E, \text{right})$
Initial:	0	0	0	0
Transition 1: $(s = D, a = \text{right}, r = -1, s' = E)$		-0.5		
Transition 2: $(s = E, a = \text{right}, r = +2, s' = F)$				1.0
Transition 3: $(s = E, a = \text{left}, r = 0, s' = D)$				
Transition 4: $(s = D, a = \text{right}, r = -1, s' = E)$		-0.25		

The agent is still at the water park MDP, but now we're going to use function approximation to represent Q-values. Recall that a policy π is greedy with respect to a set of Q-values as long as $\forall a, s \; Q(s, \pi(s)) \geq Q(s, a)$ (so ties may be broken in any way).



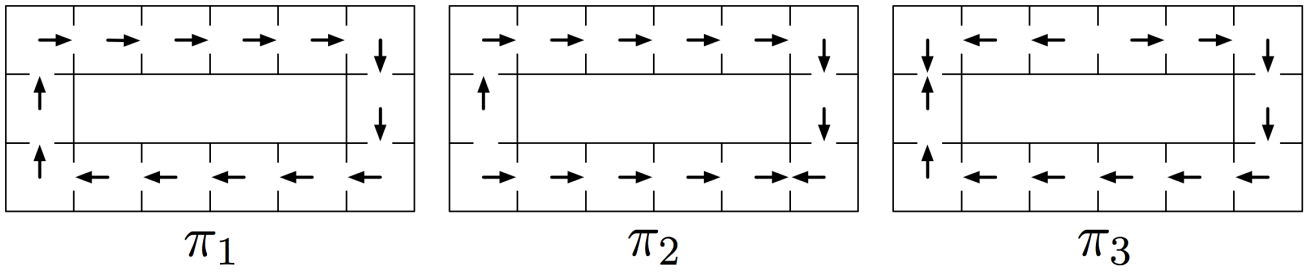
For the next subproblem, consider the following feature functions:

$$f(s,a) = \begin{cases} 1 & \text{if } a = \text{right}, \\ 0 & \text{otherwise.} \end{cases}$$

$$f'(s,a) = \begin{cases} 1 & \text{if } (a = \text{right}) \wedge \text{isSlide}(s), \\ 0 & \text{otherwise.} \end{cases}$$

(Note: isSlide(s) is true iff the state s is a slide square, i.e. either F or G.)

Also consider the following policies:



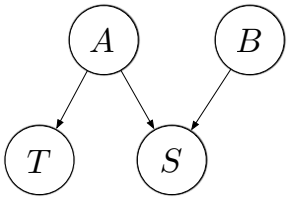
(d) [4 pts] Which are greedy policies with respect to the Q-value approximation function obtained by running the single Q-update for the transition $(s = F, a = \text{right}, r = +2, s' = G)$ while using the specified feature function? You may assume that all feature weights are zero before the update. Use discount $\gamma = 1.0$ and learning rate $\alpha = 1.0$. Circle all that apply.

f	π_1	π_2	π_3
f'	π_1	π_2	π_3

Q3. [17 pts] Bayes' Nets

Suppose that a patient can have a symptom (S) that can be caused by two different diseases (A and B). Disease A is much rarer, but there is a test T that tests for the presence of A. The Bayes' Net and corresponding conditional probability tables for this situation are shown below. For each part, you may leave your answer as a fraction.

$\mathbb{P}(A)$	
$+a$	0.1
$-a$	0.9



$\mathbb{P}(B)$	
$+b$	0.5
$-b$	0.5

$\mathbb{P}(T\|A)$		
$+a$	$+t$	1
$+a$	$-t$	0
$-a$	$+t$	0.2
$-a$	$-t$	0.8

$\mathbb{P}(S\|A,B)$			
$+a$	$+b$	$+s$	1
$+a$	$+b$	$-s$	0
$+a$	$-b$	$+s$	0.80
$+a$	$-b$	$-s$	0.20
$-a$	$+b$	$+s$	1
$-a$	$+b$	$-s$	0
$-a$	$-b$	$+s$	0
$-a$	$-b$	$-s$	1

(a) [1 pt] Compute the following entry from the joint distribution:

$\mathbb{P}(-a, -t, +b, +s) =$

$\mathbb{P}(-t|-a)\mathbb{P}(-a)\mathbb{P}(+s|+b,-a)\mathbb{P}(+b) = (0.8)(0.9)(1)(0.5) = 0.36$

(b) [2 pts] What is the probability that a patient has disease A given that they have disease B?

$\mathbb{P}(+a|+b) = \mathbb{P}(+a) = 0.1$

(c) [2 pts] What is the probability that a patient has disease A given that they have symptom S, disease B, and test T returns positive?

$$\mathbb{P}(+a|+t,+s,+b) = \frac{\mathbb{P}(+a,+t,+s,+b)}{\mathbb{P}(+t,+s,+b)} = \frac{\mathbb{P}(+a)\mathbb{P}(+t|+a)\mathbb{P}(+s|+a,+b)\mathbb{P}(+b)}{\sum_{a\in\{+a,-a\}}\mathbb{P}(a,+t,+s,+b)}$$
$$= \frac{(0.1)(1)(1)(0.5)}{(0.1)(1)(1)(0.5)+(0.9)(0.2)(1)(0.5)} = \frac{.05}{.05+.09} = \frac{5}{14} \approx 0.357$$

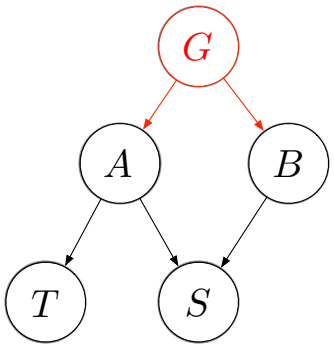
(d) [3 pts] What is the probability that a patient has disease A given that they have symptom S and test T returns positive?

$$\mathbb{P}(+a|+t,+s) = \frac{\mathbb{P}(+a,+t,+s)}{\mathbb{P}(+t,+s)} \frac{\sum_{b\in\{+b,-b\}}\mathbb{P}(+a)\mathbb{P}(+t|+a)\mathbb{P}(+s|+a,b)\mathbb{P}(b)}{\sum_{a\in\{+a,-a\}}\sum_{b\in\{+b,-b\}}\mathbb{P}(a)\mathbb{P}(+t|a)\mathbb{P}(+s|a,b)\mathbb{P}(b)} = 0.5$$

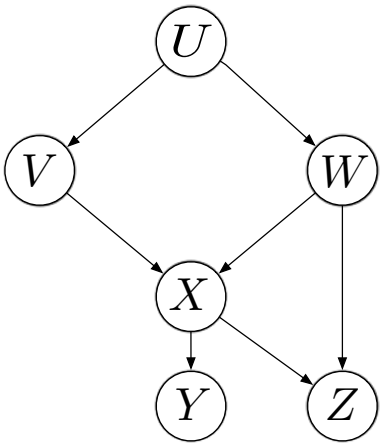
(e) [3 pts] Suppose that both diseases A and B become more likely as a person ages. Add any necessary variables and/or arcs to the Bayes' net to represent this change. For any variables you add, briefly (one sentence or less) state what they represent. Also, state one independence or conditional independence assertion that is **removed** due to your changes.

New variable(s) (and brief definition):

Removed conditional independence assertion:
There are a few. The simplest is $A \perp\!\!\!\perp B$ is no longer guaranteed. Another is $B \perp\!\!\!\perp T$.



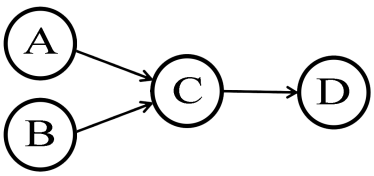
(f) [6 pts] Based only on the structure of the (new) Bayes' Net given below, circle whether the following conditional independence assertions are guaranteed to be true, guaranteed to be false, or cannot be determined by the structure alone.



1. $V \perp\!\!\!\perp W$	Guaranteed true	Cannot be determined	Guaranteed false
2. $V \perp\!\!\!\perp W \mid U$	Guaranteed true	Cannot be determined	Guaranteed false
3. $V \perp\!\!\!\perp W \mid U, Y$	Guaranteed true	Cannot be determined	Guaranteed false
4. $V \perp\!\!\!\perp Z \mid U, X$	Guaranteed true	Cannot be determined	Guaranteed false
5. $X \perp\!\!\!\perp Z \mid W$	Guaranteed true	Cannot be determined	Guaranteed false

Q4. [9 pts] Sampling

Assume the following Bayes net, and the corresponding distributions over the variables in the Bayes net:



A	$\mathbb{P}(A)$
$+a$	1/5
$-a$	4/5

B	$\mathbb{P}(B)$
$+b$	1/3
$-b$	2/3

A	B	C	$\mathbb{P}(C\|A,B)$
$+a$	$+b$	$+c$	0
$+a$	$+b$	$-c$	1
$+a$	$-b$	$+c$	0
$+a$	$-b$	$-c$	1
$-a$	$+b$	$+c$	2/5
$-a$	$+b$	$-c$	3/5
$-a$	$-b$	$+c$	1/3
$-a$	$-b$	$-c$	2/3

C	D	$\mathbb{P}(D\|C)$
$+c$	$+d$	1/2
$+c$	$-d$	1/2
$-c$	$+d$	1/4
$-c$	$-d$	3/4

(a) [1 pt] Your task is now to estimate $\mathbb{P}(+c|-a,-b,-d)$ using rejection sampling. Below are some samples that have been produced by prior sampling (that is, the rejection stage in rejection sampling hasn't happened yet). Cross out the samples that would be rejected by rejection sampling:

~~$-a \quad -b \quad +c \quad +d$~~
~~$+a \quad -b \quad -c \quad +d$~~
$-a \quad -b \quad +c \quad -d$

~~$+a \quad -b \quad -c \quad -d$~~
~~$-a \quad +b \quad +c \quad +d$~~
$-a \quad -b \quad -c \quad -d$

(b) [2 pts] Using those samples, what value would you estimate for $\mathbb{P}(+c|-a,-b,-d)$ using rejection sampling?

1/2

(c) [3 pts] Using the following samples (which were generated using likelihood weighting), estimate $\mathbb{P}(+c \mid -a,-b,-d)$ using likelihood weighting, or state why it cannot be computed.

$-a \quad -b \quad -c \quad -d$
$-a \quad -b \quad +c \quad -d$
$-a \quad -b \quad +c \quad -d$

We compute the weights of each solution, which are the product of the probabilities of the evidence variables conditioned on their parents.

$$w_1 = P(-a)P(-b)P(-d \mid -c) = 4/5 * 2/3 * 3/4$$
$$w_2 = w_3 = P(-a)P(-b)P(-d \mid +c) = 4/5 * 2/3 * 1/2$$

so normalizing, we have $(w_2 + w_3)/(w_1 + w_2 + w_3) = 4/7$.

(d) [3 pts] Below are three sequences of samples. Circle **any** sequence that could have been generated by Gibbs sampling.

Sequence 1				
1 :	$-a$	$-b$	$-c$	$+d$
2 :	$-a$	$-b$	$-c$	$+d$
3 :	$-a$	$-b$	$+c$	$+d$

Sequence 2				
1 :	$-a$	$-b$	$-c$	$+d$
2 :	$-a$	$-b$	$-c$	$-d$
3 :	$-a$	$-b$	$+c$	$+d$

Sequence 3				
1 :	$-a$	$-b$	$-c$	$+d$
2 :	$-a$	$-b$	$-c$	$-d$
3 :	$-a$	$+b$	$-c$	$-d$

The first and third sequences have at most one variable change per row, and hence could have been generated from Gibss sampling. In sequence 2, the second and third samples have both C and D changing.

Q5. [13 pts] Multiple-choice and short-answer questions

In the following problems please choose **all** the answers that apply, if any. You may circle more than one answer. You may also circle no answers (none of the above)

(a) [2 pts] Value iteration:

(i) Is a model-free method for finding optimal policies.

(ii) Is sensitive to local optima.

(iii) Is tedious to do by hand.

(iv) Is guaranteed to converge when the discount factor satisfies $0 < \gamma < 1$.

(iii) and (iv). Value iteration requires a model (an specified MDP), and is not sensitive to getting stuck in local optima.

(b) [2 pts] Bayes nets:

(i) Have an associated directed, acyclic graph.

(ii) Encode conditional independence assertions among random variables.

(iii) Generally require less storage than the full joint distribution.

(iv) Make the assumption that all parents of a single child are independent given the child.

(i), (ii), and (iii) – all three are true statements. (iv) is false – given the child, the parents are not independent.

(c) [2 pts] If we use an ϵ-greedy exploration policy with Q-learning, the estimates Q_t are guaranteed to converge to Q^* only if:

(i) ϵ goes to zero as t goes to infinity, or

(ii) the learning rate α goes to zero as t goes to infinity, or

(iii) both α and ϵ go to zero.

(ii). The learning rate must approach 0 as $t \to \infty$ in order for convergence to be guaranteed. Note that Q-learning learns off policy (in other words, it learns about the optimal policy, even if the policy being executed is sub-opitimal). This means that ϵ need not approach zero for convergence.

(d) [2 pts] True or false? Suppose X and Y are correlated random variables. Then

$$P(X = x, Y = y) = P(X = x)P(Y = y|X = x)$$

True. This is the product rule.

(e) [3 pts] **MDPs** For this question, assume that the MDP has a finite number of states.

(i) [*true* or *false*] For an MDP (S, A, T, γ, R) if we only change the reward function R the optimal policy is guaranteed to remain the same.

(ii) [*true* or *false*] Value iteration is guaranteed to converge if the discount factor (γ) satisfies $0 < \gamma < 1$.

(iii) [*true* or *false*] Policies found by value iteration are superior to policies found by policy iteration.

(f) [2 pts] **Reinforcement Learning**

(i) [*true* or *false*] Q-learning can learn the optimal Q-function Q^* without ever executing the optimal policy.

(ii) [*true* or *false*] If an MDP has a transition model T that assigns non-zero probability for all triples $T(s, a, s')$ then Q-learning will fail.