CS188: Artificial Intelligence, Spring 2009
Written Assignment 2: MDPs and Bayes Nets

Due: March 12
You can work on this in groups, but everyone should turn in his/her own work.
Don't forget your **name and login**.

1 Question 1: More Bowls of Fruit (8 points)

A researcher picks fruit from a bowl of *A*pples and *L*emons. At each state, she has either an *A*pple or *L*emon in her hand. The only action, *T*rade, places what's currently in her hand into the bowl, shakes the bowl vigorously, and removes a piece of fruit F. Let $P(F = Apple) = 0.6$, regardless of previous fruits removed.

States: A, L (A means the researcher is holding an *A*pple)
Actions: T (T is the *T*rade action)
There are no terminal states; start state is A
The researcher likes variety. Let $R(A, T, L) = 2$ and $R(L, T, A) = 3$, while all other rewards are 0.
Assume a discount rate $\lambda = 0.5$.

a) Run value iteration for this MDP for three iterations and fill in the value estimates in the table below.

	$i = 0$	$i = 1$	$i = 2$	$i = 3$
$V_i^*(A)$	0			
$V_i^*(L)$	0			

b) What are $V^*(A)$ and $V^*(L)$ in this MDP? *Hint: write down the Bellman equations and solve them.*

c) The researcher considers one more action: when holding a *L*emon, she can *S*queeze it, which makes a mess. $T(L, S, L) = 1$ and $R(L, S, L) = -1$ (S is the *S*queeze action). All previous rewards and transitions still apply. What is $Q^*(L, S)$? *Hint: you can use your answer from (b)*

d) We call an MDP zeroth-order if $T(s_1, a_1, s') = T(s_2, a_2, s')$ for all states s_1, s_2 and s' and all actions a_1 and a_2. That is, the successor state does not depend on the current state or the action taken (like the original bowl of fruit problem). If $R(s, a, s')$ is between $-k$ and k inclusive for all (s, a, s'), what is the maximum difference $|V^\pi(s_1) - V^\pi(s_2)|$ in a zeroth-order MDP for any pair of states s_1 and s_2 under any policy π?

2 Question 2: Q-Earning (6 points)

An enterprising 188 student builds a q-learning agent to buy and sell real estate. The agent begins owning *N*othing. When the agent owns *N*othing, it can try to buy land (*BuyL*) or buy a mansion (*BuyM*). If successful, the agent then owns *L*and or a *M*ansion. When the agent owns property, it can only try to *S*ell. Once the property is sold, the scenario ends in the *T*erminal state. Assume a discount factor of $\lambda = 1$.

a) At first, the student makes decisions for two episodes and the agent just learns $Q(s, a)$. Fill in the table with Q-value estimates after each observation. Use a learning rate of $\alpha = 0.5$. All estimates begin at 0. Terminal state T has no actions and a value of 0.

Observation				**New Q Estimates**			
State s	**Action** a	**Successor** s'	**Reward** r	$Q(N, BuyL)$	$Q(N, BuyM)$	$Q(L, S)$	$Q(M, S)$
			Initial values:	0	0	0	0
N	$BuyL$	L	-10				
L	Sell	T	20				
N	$BuyL$	N	-2				
N	$BuyM$	M	-20				
M	Sell	M	-8				
M	Sell	T	60				

b) What is the optimal policy from this q-learning agent, and what is the q-learning agent's estimate of the expected sum of discounted future rewards starting in N under this policy?

c) If we instead use these observations to estimate a transition model, what would it be?

$T(N, BuyL, N) =$		$T(L, Sell, L) =$	
$T(N, BuyL, L) =$		$T(L, Sell, T) =$	
$T(N, BuyM, N) =$		$T(M, Sell, M) =$	
$T(N, BuyM, M) =$		$T(M, Sell, T) =$	

d) If the same sequence of 6 observations in (a) repeated indefinitely and the q-learner very slowly decreased its learning rate, what would $Q^*(N, BuyM)$ converge to? Justify your answer.

3 Question 3: Dice on Ice (7 points)

Ms. Pacman has special 6-sided dice that are fair at room temperature, but tend to come up 6 when frozen. Frozen dice come up 6 half the time, and come up 1 through 5 with equal probability the other half. She offers you a simple game: pay $1 and roll two dice. If you roll 11 or 12, you get $10.

(a) Rolling two fair dice (equal probability of numbers 1 through 6), what is your expected payoff.

(b) Rolling two dice, one fair and one frozen, what is your expected payoff?

(c) Rolling two frozen dice what is your expected payoff?

(d) You pick two dice randomly: each one is frozen with probability $\frac{1}{2}$. What is your expected payoff?

(e) You picked dice as in (d), then rolled a twelve. What is the probability that both dice were frozen?

4 Question 4: Studious Students (9 points)

A student may or may go to class (C), may or may not ace the exam ($A = a$), and may or may not beat the curve ($B = b$). The possible outcomes are listed in the following joint probability table:

A	B	C	$P(A, B, C)$
a	b	c	0.280
a	$\neg b$	c	0.120
$\neg a$	b	c	0.280
$\neg a$	$\neg b$	c	0.120
a	b	$\neg c$	0.016
a	$\neg b$	$\neg c$	0.024
$\neg a$	b	$\neg c$	0.064
$\neg a$	$\neg b$	$\neg c$	0.096

(a) What is the distribution $P(A, B)$? Fill in the table below.

A	B	$P(A, B)$
a	b	
a	$\neg b$	
$\neg a$	b	
$\neg a$	$\neg b$	

(b) Are A and B independent? Justify your answer using the actual probabilities computed in part (a).

(c) What is the marginal distribution $P(C)$?.

C	P(C)
c	
$\neg c$	

(d) What is the posterior distribution over C given that $B = b$?

B	C	$P(C\|B = b)$
b	c	
b	$\neg c$	

(e) What is the posterior distribution over C given that $A = a$ and $B = b$?

A	B	C	$P(C\|A = a, B = b)$
a	b	c	
a	b	$\neg c$	

(f) Briefly explain why the pattern amongst $P(C)$, $P(C|B = b)$, and $P(C|A = a, B = b)$ makes sense.

A	B	C	$P(A, B, C)$
a	b	c	0.280
a	$\neg b$	c	0.120
$\neg a$	b	c	0.280
$\neg a$	$\neg b$	c	0.120
a	b	$\neg c$	0.016
a	$\neg b$	$\neg c$	0.024
$\neg a$	b	$\neg c$	0.064
$\neg a$	$\neg b$	$\neg c$	0.096

Repeated for convenience

(g) What is $P(A|C)$?

A	C	$P(A|C)$
a	c	
$\neg a$	c	
a	$\neg c$	
$\neg a$	$\neg c$	

(h) Is A conditionally independent of B given C? Justify your answer using the probabilities.

(i) Draw the Bayes net structure with the fewest arcs that can express this distribution.