ASEN 5519-003 Decision Making under Uncertainty
Quiz 2: RL, POMDPs, Bayes Nets, and Games

Show all work and box answers.
You may consult any source, but you may NOT communicate with any person except the instructor.

Question 1. (5 pts) The update step for SARSA can be written as

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma Q(s',a') - Q(s,a))$$

Using the same notation, write the update step for (a) **Q-learning** and (b) **double Q-learning**.

Question 2. (15 pts) Consider a 1 state, 2 action infinite horizon MDP defined as follows:

$$\mathcal{S} = \{1\} \quad \mathcal{A} = \{1,2\} \tag{1}$$

$$\mathcal{R}(1,1) = 1 \quad \mathcal{R}(1,2) = 0 \tag{2}$$

$$\mathcal{T}(1 \mid 1,a) = 1 \; \forall \, a \quad \gamma = 0.9 \tag{3}$$

After running the Q learning algorithm for a long time until it converges,

a) What numerical values will the state-action (Q) value function converge to?

b) After convergence, how often would a softmax policy with $\lambda = 1$ choose action 2?

c) After convergence, how often would an epsilon greedy policy with $\epsilon = 0.1$ choose action 2?

Question 3. (30 pts) Consider the following modified Tiger POMDP with a modified action space and rewards:

$$\mathcal{S} = \mathcal{O} = \{TL, TR\} \tag{4}$$

$$\mathcal{A} = \{OL, OR, L, J\} \quad (J \text{ corresponds to jumping up and down}) \tag{5}$$

$$\mathcal{R}(s,a) = \begin{cases} -1 \text{ if } a = J \quad \text{(penalty for physical exertion)} \\ -15 \text{ if } s = TL \text{ and } a = OL \text{ or } s = TR \text{ and } a = OR \quad \text{(opening tiger door)} \\ +5 \text{ if } s = TL \text{ and } a = OR \text{ or } s = TR \text{ and } a = OL \quad \text{(opening other door)} \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

$$\mathcal{T}(s' \mid s,a) = \begin{cases} 0.5 \text{ if } a \in \{OL, OR\} \quad \text{(randomly reset after opening door)} \\ 1.0 \text{ if } s' = s \text{ and } a \in \{L, J\} \quad \text{(stay the same)} \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

$$\mathcal{Z}(o \mid a,s') = \begin{cases} 0.5 \text{ if } a \in \{OL, OR, J\} \\ 0.85 \text{ if } a = L \text{ and } o = s' \\ 0.15 \text{ if } a = L \text{ and } o \neq s' \end{cases} \tag{8}$$

$$\gamma = 0.95 \tag{9}$$

a) Calculate the and write out **one step** alpha vectors for each action.

b) Draw the **one step** alpha vectors in the manner done in class.

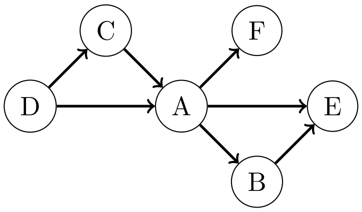
c) Suppose that you start with an initial belief of $b(TR) = 0.6$. According to the policy encoded in the alpha vectors above, what action should you take?

d) Suppose that you take the action above, and receive the observation TR. What is the new belief?

e) According to the policy encoded in the alpha vectors above, which action should you take after this belief update?

f) According to the policy encoded in the alpha vectors above, under what circumstances would you take the J action?

Question 4. (25 pts) Consider the following Bayesian network structure:



a) True or False: $F \perp B \mid E$? Justify your answer.

b) True or False: $B \perp D \mid A$? Justify your answer.

c) Consider the following claim: "If $D = 1$ and $A = 2$, then B is always 3, but if $D = 4$ and $A = 2$, then B is always 1." Can this claim be proved or disproved with the Bayes net structure above? Explain.

Question 5. (25 pts)

a) Give an example of a simple game with 2 players with 2 actions each that has at least 4 pure strategy Nash equilibria.

b) Give an example of a simple game with 2 players with 3 actions each that has exactly 3 pure strategy Nash equilibria.

c) What are the pure Nash equilibria for the following game?

		Player 2		
		a	b	c
Player 1	a	4,4	2,5	0,0
	b	5,2	3,3	0,0
	c	0,0	0,0	10,10