Last Name: ______________________ Fist Name: ____________ SID#:___________
Collaborators:____

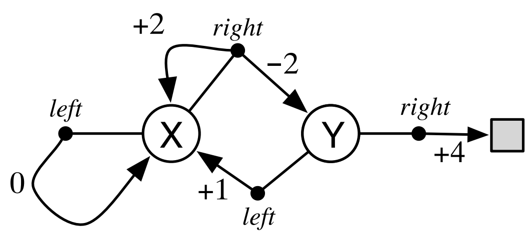
CMPUT 609 Written 2: Markov Decision Processes

Due: Thursday Sept 22 in Gradescope by 2pm (no slip days)

Policy: Can be solved in groups (acknowledge collaborators) must be written up individually. There are a total of 100 points on this assignment.

Be sure to explicitly answer each subquestion posed in each exercise. The boxes are in an appropriate size to contain a long answer. If you think you need more than the space allocated, there is a better way to respond the posed questions.

Question 1 [15 points]: *Trajectories, returns, and values.*



Consider the MDP above, in which there are two states, X and Y, two actions, *right* and *left*, and the deterministic rewards on each transition are as indicated by the numbers. Note that if action *right* is taken in state X, then the transition may be either to X with a reward of $+2$ or to Y with a reward of -2. These two possibilities occur with probabilities 2/3 (for the transition to X) and 1/3 (for the transition to state Y).
Consider two deterministic policies, π_1 and π_2:

$$\pi_1(\mathsf{X}) = \textit{left} \qquad \pi_2(\mathsf{X}) = \textit{right}$$
$$\pi_1(\mathsf{Y}) = \textit{right} \qquad \pi_2(\mathsf{Y}) = \textit{right}$$

(a) (2 pts.) Show a typical trajectory (sequence of states, actions and rewards) from X for policy π_1:

(b) (2 pts.) Show a typical trajectory (sequence of states, actions and rewards) from X for policy π_2:

(c) (2 pts.) Assuming the discount-rate parameter is $\gamma = 0.5$, what is the return from the initial state for the second trajectory?

(d) (2 pts.) Assuming $\gamma = 0.5$, what is the value of state Y under policy π_1?

(e) (2 pts.) Assuming $\gamma = 0.5$, what is the action-value of X,*left* under policy π_1?

(f) (5 pts) Assuming $\gamma = 0.5$, what is the value of state X under policy π_2?

Question 2 [6 points - 3 for each subquestion]: *Problem with maze running.*
See Exercise 3.5 in the SB textbook.

Question 3 [5 points]: *Broken vision system.*
See Exercise 3.6 in the SB textbook, second edition.

Question 4 [16 points]: *Bellman equation for action values, q_π.*
See Exercise 3.7 in the SB textbook, second edition.

Question 5 [14 points, 7 for each subquestion]: *Verify Bellman equation in gridworld example* (This differs from the texbook).
The Bellman equation (3.12) must hold for each state for the value function v_π shown in Figure 3.5 (right figure, see SB textbook, second edition). As an example, show numerically that this equation holds for the state just below the center state, valued at -0.4, with respect to its four neighboring states, valued at $+0.7$, -0.6, -1.2, and -0.4. Also, show numerically that this equation also holds for the state B (depicted in Figure 3.5, left), valued at $+5.3$ (These numbers are accurate only to one decimal place).

Question 6 [14 points, 4 for each subquestion, 6 for the proof]: *Adding a constant reward in a continuing task.*
See Exercise 3.9 in the SB textbook, second edition.

Question 7 [9 points, 3 for each subquestion, 3 for the example]: *Adding a constant reward in an episodic task.*
See Exercise 3.10 in the SB textbook, second edition.

Question 8 [8 points, 4 for each equation]: *Half-backup* v_π.
See Exercise 3.11 in the SB textbook, second edition.

Question 9 [8 points, 4 for each equation]: *Half-backup* q_π.
See Exercise 3.12 in the SB textbook, second edition.

Question 10 [5 points]: *Changes in the optimal policy.*
Suppose we have two problems with the same state and action spaces. Let the optimal action-value functions of the two problems be denoted q_* and q'_*, and suppose it happens to be the case that $q'_*(s, a) = q_*(s, a) + f(s), \forall s, a$ for some function $f : \mathcal{S} \mapsto \Re$. What is the relationship between the optimal policies π_* and π'_* for the two problems?