Assignment 1

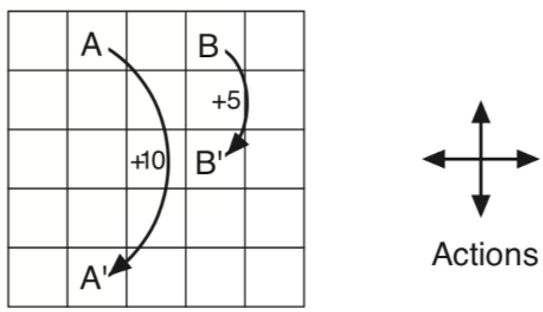
Reinforcement Learning

CISC 474, Fall, 2022

Due Monday, October17, 2022 before Midnight

Markov Decision Processes (90 pts)

In class we discussed the problem of how to deal with a *markovian* environment. Consider the following grid:



In it, for all the states (cells in the grid), each one of the actions (north, south, east, and west) is chosen with probability ¼. The agent then moves with probability 1 to the chosen direction. While moving, if the agent hits a wall, it cannot move and it receives a reward of -1; if moving to a cell in the grid, the reward is 0.; if it reaches cell A (1,2), it moves to cell A' (5,2) and receives a reward +10; and, if it reaches cell B (1,4) it moves to cell B' (3,4) and receives a reward +5.

You are required to write **2** programs (in Python 3) to find the state-value for each one of the states for discount rates of 0.85 and 0.75. You are also required to find the state-value for each state when the grid is 7x7 and A is located in (3,2), A' in (7,2), B in (1,6) and B' in (4,6). The marking scheme is:

- 25% for finding the correct state-value table for the 5x5 case.
- 25% for finding the correct state-value table for the 7x7 case.
- 25% for your code. You should be able to reuse your solution for the 5x5 case in your 7x7 case.
- 25% for your report. You should clearly identify your solution approach and show your results.

The 2 programs you will write are:

1. A linear solver to solve the system $A\bar{x} = \bar{b}$.
2. A dynamic programming approach (policy iteration or value iteration).

Please, notice that your code will be checked by a system similar to Turnitin for plagiarism.

Hint: you can check your solution by running the problem presented in the lectures (example 3.5 in the book). In that case, $\gamma = 0.9$.

State-value and Action-value functions (10 pts)

Recall that

$$v_\pi(s) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,|S_t = s\right]$$

$$q_\pi(s,a) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,|S_t = s, A_t = a\right]$$

Using backup diagrams (or mathematical derivations):

1. Write v_π in terms of q_π. (5 pts)
2. Write in q_π terms of v_π. (5 pts)