Homework 2: Finite MDPs and Bellman Equations

Due: Monday, September 27th 11:59 pm

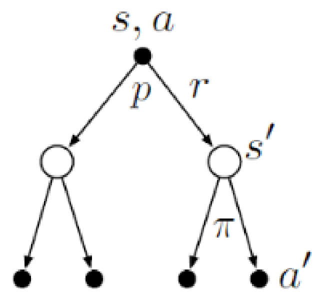
Q1. Prove that adding a constant c to all rewards adds a constant (V_c) to the value of all states, and thus does not affect the relative values of any states under any policies. What is V_c in terms of c and γ. Hint: start from $G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$ and calculate $V_\pi(s)$.

Q2. Suppose $\gamma = 0.5$ and the following sequence of rewards is received

$$R_1 = -1, R_2 = 2, R_3 = 6, R_4 = 3, R_5 = 2,$$

with $T = 5$. What are $G_0, G_1, \ldots, G_5$? Hint: Work backwards.

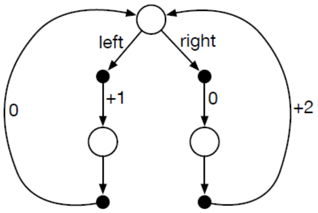
Q3. We have derived Bellman equation for $V_\pi(s)$ that expresses the relationship between the value of a state and the values of its successor states (backup diagram included in the slides). What is the Bellman equation for $Q_\pi(s,a)$? It must give action value $Q_\pi(s,a)$ in terms of $Q_\pi(s',a')$. Use the following backup diagram to write the equation and explain its individual terms.



q_π backup diagram

Q4. Consider the continuing MDP shown below. The only decision to be made is that in the top state, where two actions are available, **left** and **right**. The numbers show the rewards that are received deterministically after each action. There are exactly two deterministic policies, π_L, π_R. What policy is optimal if

a. $\gamma = 0$
b. $\gamma = 0.9$
c. $\gamma = 0.5$



Explain your computation for each case.

Evaluation: we will grade your submission according to the following table:

Item	**COMP4600**	**COMP5300**
Question 1	20	20
Question 2	20	20
Question 3	30	30
Question 4	30	30

Note 1: The parts marked with **(*)** are optional for COMP4600 (undergraduates) but mandatory for COMP5300 (graduates). This homework does not include any optional section.

Note 2: All explanations, formulae, and answers should be included in a single Jupyter Notebook (`.ipynb`) file. Include your name as part of the filename and submit through Blackboard.

Submission: By 11:59pm on Monday, September 27th 2021, submit your `student_name.ipynb` files on Blackboard. Make sure everything is entirely contained within this file and it runs without any error.

Late Policy: Up to two late days are allowed, but a grade penalty of 50% and 75% will be applied at the first and second day, respectively.