CS 747 (Spring 2025) Week 4 Test (Batch 1)

5.35 p.m. – 6.00 p.m., February 6, 2025, LA 001

Name: _____ Roll number: _____

Note. There is one question in this test. You can use the space on both pages for your answer. Draw a line (either vertical or horizontal) and do all your rough work on one side of it.

Question 1. Your task is to construct a family of 2-state, 2-action MDPs that all have the same sets of states and actions, same transition and reward functions, but which differ in their discount factors. The family needs to satisfy a particular constraint, which is specified below.

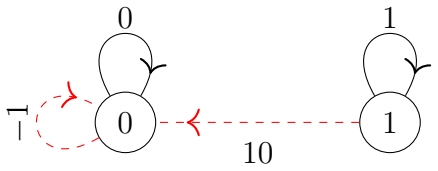
Suppose each MDP in the family has a set of states $S = \{0, 1\}$ and a set of actions is $A = \{0, 1\}$. Define $\gamma_0 = 0.9$. You must specify transition function T and reward function R such that if $\pi : S \to A$ is an optimal policy for the MDP (S, A, T, R, γ) for any fixed $\gamma \in (0, \gamma_0)$, then π is *not* an optimal policy for any MDP (S, A, T, R, γ'), where $\gamma' \in (\gamma_0, 1)$. In other words, T and R must be such that no two MDPs with discount factors on different sides of $\gamma_0 = 0.9$ have a common optimal policy.

Fill out the entries of T and R in the table below with numeric values (specific numbers, rather than variables). There are multiple satisfying solutions; all you need to do is specify one, and provide an explanation as to why it satisfies the required property. You might find it convenient to draw a state-transition diagram—but make sure to also fill out the table. [3 marks]

s	a	s'	$T(s, a, s')$	$R(s, a, s')$
0	0	0		
0	0	1		
0	1	0		
0	1	1		
1	0	0		
1	0	1		
1	1	0		
1	1	1		

Answer 1.

For questions of this nature, it is a good idea to begin with "simple, convenient" MDPs, and add complexity to them only if required. For this particular problem, let us only consider MDPs with deterministic transitions. Moreover, we can configure one state to always (that is, independent of γ) have the same optimal action, thereby leaving us only the other state to worry about. From this other state, we require one policy to dominate for $\gamma < \gamma_0$, and another one to dominate for $\gamma > \gamma_0$. The following MDP family achieves this effect. All transitions shown have probability 1; annotations show the reward. Transitions of action 0 are shown in black (solid) and those of action 1 in red (dashed).



Notice that for all $\gamma \in (0, 1)$, every policy that takes action 0 at state 0 has a value of 0 there, while every policy that takes action 1 at state 0 has value $-\frac{1}{1-\gamma}$ there. Thus, an optimal policy must necessarily take action 0 at state 0. So we fix action 0 at state 0 and consider state 1: that is, we consider policies $\pi_1 = 00$ and policy $\pi_2 = 01$. Clearly, $V^{\pi_1}(1) = \frac{1}{1-\gamma}$ and $V^{\pi_2}(1) = 10$. If $\gamma < \gamma_0$, π_2 dominates π_1, whereas if $\gamma > \gamma_0$, π_1 dominates π_2. If $\gamma = \gamma_0$, then both π_1 and π_2 are optimal.

Here is the table corresponding to our MDP family.

s	a	s'	$T(s, a, s')$	$R(s, a, s')$
0	0	0	1	0
0	0	1	0	0
0	1	0	1	-1
0	1	1	0	0
1	0	0	0	0
1	0	1	1	1
1	1	0	1	10
1	1	1	0	0

CS 747 (Spring 2025) Week 4 Test (Batch 2)

6.15 p.m. – 6.40 p.m., February 6, 2025, LA 001

Name: _____ Roll number: _____

Note. There is one question in this test. You can use the space on both pages for your answer. Draw a line (either vertical or horizontal) and do all your rough work on one side of it.

Question 1. This question requires you to construct a 2-state, 2-action MDPs that has a pair of "incomparable" policies. Name the set of states $S = \{0, 1\}$ and a set of actions is $A = \{0, 1\}$. You must specify transition function T, reward function R, and discount factor γ so that the following property is achieved by your MDP:

> *There exist policies* $\pi_1 : S \to A$ *and* $\pi_2 : S \to A$ *such that* $V^{\pi_1}(0) > V^{\pi_2}(0)$ *and* $V^{\pi_1}(1) < V^{\pi_2}(1)$.

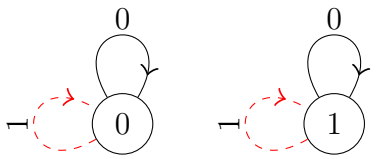
Fill out the entries of T, R, and γ in the table below with numeric values (specific numbers, rather than variables). Specify the relevant policies and their values in the accompanying table. The discount factor must be strictly positive: that is, $\gamma \in (0, 1)$. There are multiple satisfying solutions; all you need to do is specify one. You might find it convenient to draw a state-transition diagram—but make sure to also fill out the tables. [3 marks]

s	a	s'	$T(s, a, s')$	$R(s, a, s')$
0	0	0		
0	0	1		
0	1	0		
0	1	1		
1	0	0		
1	0	1		
1	1	0		
1	1	1		
$\gamma =$				

s	$\pi_1(s)$	$\pi_2(s)$	$V^{\pi_1}(s)$	$V^{\pi_2}(s)$
0				
1				

Answer 1.

For questions of this nature, it is a good idea to begin with "simple, convenient" MDPs, and add complexity to them only if required. For this particular problem, let us only consider MDPs with deterministic transitions. In fact, it is even more convenient to consider MDPs in which each state transitions only into itself, with the action determining only the reward. Drawn below is an MDP that satisfies the required property for all $\gamma \in (0, 1)$. All transitions shown have probability 1; annotations show the reward. Transitions of action 0 are shown in black (solid) and those of action 1 in red (dashed).



Take π_1 as the policy 10, and π_2 as the policy 01. The policies are incomparable for all $\gamma \in (0, 1)$. Here are filled up tables for $\gamma = 0.5$.

s	a	s'	$T(s, a, s')$	$R(s, a, s')$
0	0	0	1	0
0	0	1	0	0
0	1	0	1	1
0	1	1	0	0
1	0	0	0	0
1	0	1	1	0
1	1	0	0	0
1	1	1	1	1
$\gamma = 0.5$				

s	$\pi_1(s)$	$\pi_2(s)$	$V^{\pi_1}(s)$	$V^{\pi_2}(s)$
0	1	0	2	0
1	0	1	0	2