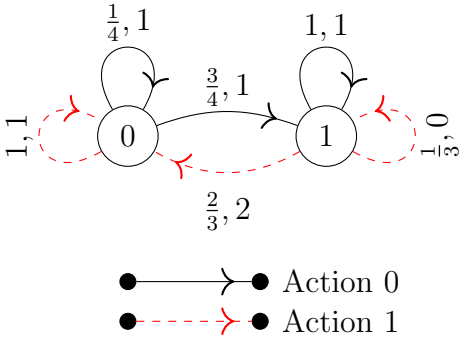
CS 747 (Spring 2025) Week 5 Test (Batch 1)

5.35 p.m. – 6.00 p.m., February 13, 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. Consider the MDP (S, A, T, R, γ) given below, with $S = \{0, 1\}$ and $A = \{0, 1\}$. The transition function T and reward function R are specified as annotations in the state-transition diagram. Arrows are annotated with "transition probability, reward" pairs; zero-probability transitions are not shown. For easy reference, T and R are also listed in the table below. The discount factor is $\gamma = \frac{2}{3}$.



s	a	s'	$T(s, a, s')$	$R(s, a, s')$
0	0	0	1/4	1
0	0	1	3/4	1
0	1	0	1	1
0	1	1	0	0
1	0	0	0	0
1	0	1	1	1
1	1	0	2/3	2
1	1	1	1/3	0

Consider a run of value iteration, initialised with the vector $V^0 = [2, -1]$, with the interpretation that $V^0(0) = 2$ and $V^0(1) = -1$. Denote the vector obtained after the first update $V^1 = B^\star(V^0)$. Compute V^1 [2 marks] and $\|V^1 - V^0\|_\infty$ [1 mark].

Answer 1.

$$
\begin{aligned}
V^1(0) &= \max\{A^0, B^0\}, \text{ where} \\
A^0 &= \frac{1}{4}(1+\gamma V^0(0)) + \frac{3}{4}(1+\gamma V^0(1)) = \frac{1}{4} + \frac{1}{4}\cdot\frac{2}{3}\cdot 2 + \frac{3}{4} + \frac{3}{4}\cdot\frac{2}{3}\cdot(-1) = \frac{5}{6}; \\
B^0 &= 1(1+\gamma V^0(0)) = \frac{7}{3}.
\end{aligned}
$$

Therefore, $V^1(0) = \frac{7}{3}$.

$$
\begin{aligned}
V^1(1) &= \max\{A^1, B^1\}, \text{ where} \\
A^1 &= 1(1+\gamma V_0(1)) = 1 + \frac{2}{3}\cdot(-1) = \frac{1}{3}; \\
B^1 &= \frac{2}{3}(2+\gamma V^0(0)) + \frac{1}{3}\gamma V^0(1) = \frac{4}{3} + \frac{8}{9} - \frac{2}{9} = 2.
\end{aligned}
$$

Therefore, $V^1(1) = 2$.

In Summary, $V^1 = [\frac{7}{3}, 2]$.

Also, $\|V^1 - V^0\|_\infty = \max\left\{|\frac{7}{3} - 2|, |2 - (-1)|\right\} = 3$.

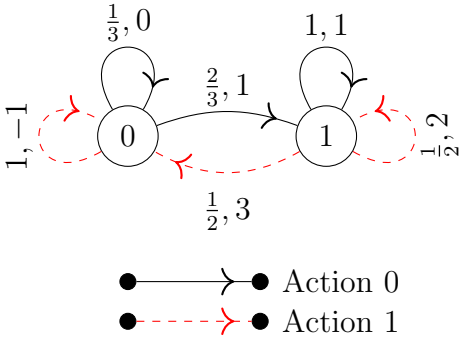
CS 747 (Spring 2025) Week 5 Test (Batch 2)

6.15 p.m. – 6.40 p.m., February 13, 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. Consider the MDP (S, A, T, R, γ) given below, with $S = \{0, 1\}$ and $A = \{0, 1\}$. The transition function T and reward function R are specified as annotations in the state-transition diagram. Arrows are annotated with "transition probability, reward" pairs; zero-probability transitions are not shown. For easy reference, T and R are also listed in the table below. The discount factor is $\gamma = \frac{3}{4}$.



s	a	s'	$T(s, a, s')$	$R(s, a, s')$
0	0	0	1/3	0
0	0	1	2/3	1
0	1	0	1	-1
0	1	1	0	0
1	0	0	0	0
1	0	1	1	1
1	1	0	1/2	3
1	1	1	1/2	2

Write down the linear program induced by this MDP, as per the formulation discussed in class. In other words, specify the variables, objective function, and constraints. Simplify any linear expressions such that the coefficients are integers; re-order inequalities so they read as "expression ≤ 0". For example, "$\frac{4}{3}x_1 - (5 \times 3)x_2 \geq 8$" must equivalently be written as "$-4x_1 + 45x_2 + 24 \leq 0$". [3 marks]

Answer 1.

We require a variable to denote the optimal value from each state; let us call these variables V_0 and V_1 for states 0 and 1, respectively.

The convention is to take the objective function (to maximise) as $(-V_0 - V_1)$, although other combinations that put a negative weight on each of the variables will also work (for example, $(-2V_0 - 43V_1)$.)

For each $s \in S, a \in A$, we have the constraint

$$V_s \geq \sum_{s' \in S} T(s, a, s')\{R(s, a, s') + \gamma V_{s'}\}.$$

For $s = 0, a = 0$, we have

$$V_0 \geq \frac{1}{3}(\gamma V_0) + \frac{2}{3}(1 + \gamma V_1) \iff V_0 \geq \frac{1}{4}V_0 + \frac{2}{3} + \frac{1}{2}V_1 \iff -9V_0 + 6V_1 + 8 \leq 0.$$

For $s = 0, a = 1$, we have

$$V_0 \geq -1 + \gamma V_0 \iff -V_0 - 4 \leq 0.$$

For $s = 1, a = 0$, we have

$$V_1 \geq 1 + \gamma V_1 \iff -V_1 + 4 \leq 0.$$

For $s = 1, a = 1$, we have

$$V_1 \geq \frac{1}{2}(3 + \gamma V_0) + \frac{1}{2}(2 + \gamma V_1) \iff V_1 \geq \frac{3}{2} + \frac{3}{8}V_0 + 1 + \frac{3}{8}V_1 \iff 3V_0 - 5V_1 + 20 \leq 0.$$

In summary, we have the following linear program with variables V_0 and V_1.

Maximise $(-V_0 - V_1)$

subject to:

$$\begin{aligned} -9V_0 + 6V_1 + 8 &\leq 0, \\ -V_0 - 4 &\leq 0, \\ -V_1 + 4 &\leq 0, \\ 3V_0 - 5V_1 + 20 &\leq 0. \end{aligned}$$