Problem Set 2

Teera Tesharojanasup

Northeastern University, Boston

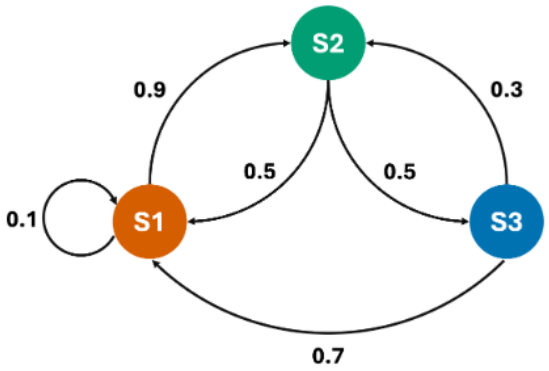
July 23rd, 2024

Overview

Problem set 2 for CS 4100 Summer II. Taught by assistant teaching professor, Rajagopal Venkat. [1]

1 Markov Models

Consider the following Markov model with 3 states, **S1**, **S2** and **S3**. The transition probabilities of the model are represented along the edges (note that each edge is directional).



Q1 **Given the above Markov model, explain whether the stationary distribution depends on the start state (without actually computing the stationary distribution).** **(2)**

Given the above Markov model, the stationary distribution does not depend on the start state. Given any start state, the probability distribution will converge towards the stationary distribution as n grows larger as long as there is a path that exists between any two states which the Markov model above satisfies.

Q2 **Given the initial probability distribution $p_0 = [0.1, 0.6, 0.3]$ for states $S1, S2$, and $S3$ respectively, compute the stationary distribution for the given Markov model. You may use a program to do the computations. Report only the final stationary distribution π.** **(2)**

$\pi = [0.38636364,\ 0.40909091,\ 0.20454545]$

Q3 **What is the probability of the following transition sequence: $S1 \to S1 \to S2 \to S3$? (Use the stationary distribution computed above, and show your calculations.)** **(2)**

Probability of starting at S1 = 0.3864

$$P(S1 \to S1 \to S2 \to S3) = (\text{Probability of starting at S1}) \cdot P(S1|S1) \cdot P(S2|S1) \cdot P(S3|S2)$$
$$P(S1 \to S1 \to S2 \to S3) = 0.3864 \cdot 0.1 \cdot 0.9 \cdot 0.5$$
$$P(S1 \to S1 \to S2 \to S3) = 0.017388$$

Q4 **Given that we start in the state $S2$, what is the probability of returning to $S2$ after two transitions? (Show your calculations.)** **(2)**

$$\text{Probability of starting at S2} = 0.409$$

$$P(S2 \to X \to S2) = P(S2 \to S1 \to S2) + P(S2 \to S2 \to S2) + P(S2 \to S3 \to S2)$$
$$P(S2 \to X \to S2) = (0.409 \cdot 0.5 \cdot 0.9) + (0.409 \cdot 0 \cdot 0) + (0.409 \cdot 0.5 \cdot 0.3)$$
$$P(S2 \to X \to S2) = 0.2454$$

Q5 **Given the starting probability distribution $p_0 = [0.1, 0.6, 0.3]$, what is the probability of ending up in $S2$ after exactly two transitions? (Show your calculations.)** **(2)**

$$\text{Probability of starting at S1} = 0.1$$
$$\text{Probability of starting at S2} = 0.6$$
$$\text{Probability of starting at S3} = 0.3$$

$$\begin{aligned} P(X \to Y \to S2) = {} & P(S1 \to S1 \to S2) + P(S1 \to S2 \to S2) + P(S1 \to S3 \to S2) \\ & + P(S2 \to S1 \to S2) + P(S2 \to S2 \to S2) + P(S2 \to S3 \to S2) \\ & + P(S3 \to S1 \to S2) + P(S3 \to S2 \to S2) + P(S3 \to S3 \to S2) \end{aligned}$$

$$\begin{aligned} = {} & (0.1 \cdot 0.1 \cdot 0.9) + (0.1 \cdot 0.9 \cdot 0) + (0.1 \cdot 0 \cdot 0.3) \\ & + (0.6 \cdot 0.5 \cdot 0.9) + (0.6 \cdot 0 \cdot 0) + (0.6 \cdot 0.5 \cdot 0.3) \\ & + (0.3 \cdot 0.7 \cdot 0.9) + (0.3 \cdot 0.3 \cdot 0) + (0.3 \cdot 0 \cdot 0.3) \end{aligned}$$

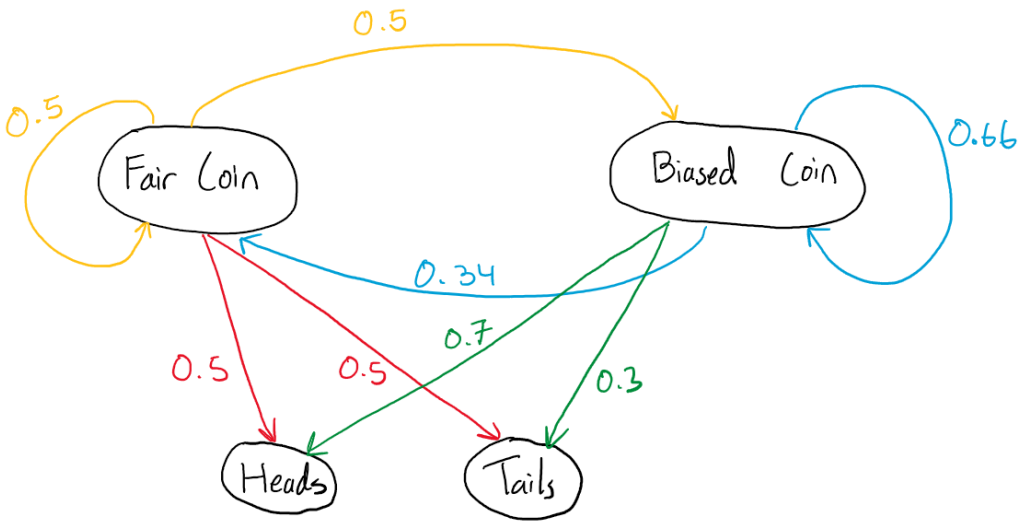
$$= 0.558$$

2 Hidden Markov Models

Recall Hidden Markov Models (HMM) from class, where we applied this approach to sequence labeling tasks such as parts-of-speech tagging or named entity recognition. Here, your task is to construct and use an HMM model to make inferences about a coin-flipping game with the following rules.

Your professor produces two identical looking coins. However, only one of the coins is a fair coin, and the other is a biased coin that produces an outcome of **Heads** 70% of the time. The professor always knows which coin is the fair one, and will perform three coin flips in total. Between each flip, the professor may swap the coin, following the rule that if a fair coin is flipped in one round, then in the next round, the professor chooses a coin completely at random. If, however, the biased coin is flipped in any round, then the professor is twice as likely to choose the biased coin again in the next round, as compared to the fair coin. As the three flips are performed, you observe the outcomes **Heads**, **Heads** and **Tails** respectively.

Q6 **Draw the HMM diagram for this game showing transitions between hidden states, and emissions to outcomes with the corresponding probabilities labeled along the edges. Hand-drawn figures accepted for this question, provided the grader can read everything clearly. (For reference, see the second diagram in our HMM notes, showing the Very Late, Late and On Time hidden states, and the Happy and Sad outcomes.)** **(5)**



Q7 **Given the observed outcome sequence, predict which coin was most likely flipped in each of the three turns (i.e., compute the most likely hidden sequence). Show all your intermediate calculations and use the stationary distribution to reason about which coin was flipped in the very first round.** **(10)**

FC = Fair Coin, BC = Biased Coin, H = Heads, T = Tails

$$\textit{Transition Matrix} = \begin{array}{c} \\ FC \\ BC \end{array} \begin{array}{c} \begin{array}{cc} FC & BC \end{array} \\ \begin{pmatrix} 0.5 & 0.5 \\ 0.34 & 0.66 \end{pmatrix} \end{array}$$

$$\textit{Observation Matrix} = \begin{array}{c} \\ H \\ T \end{array} \begin{array}{c} \begin{array}{cc} H & T \end{array} \\ \begin{pmatrix} 0.5 & 0.5 \\ 0.7 & 0.3 \end{pmatrix} \end{array}$$

$$\textit{Stationary Distribution} : \pi = [0.4047619,\ 0.5952381]$$

One Sequence Permutations :

$$\begin{array}{l} FC \\ \downarrow \quad P(Start\ FC) \cdot P(H|FC) = 0.4048 \cdot 0.5 = 0.2024 \\ H \end{array}$$

$$\begin{array}{l} BC \\ \downarrow \quad P(Start\ BC) \cdot P(H|BC) = 0.5952 \cdot 0.7 = 0.4166 \\ H \end{array}$$

Two Sequence Permutations :

$$\begin{array}{cc} FC \to FC & \\ \downarrow \quad \downarrow & 0.2024 \cdot P(FC|FC) \cdot P(H|FC) = 0.2024 \cdot 0.5 \cdot 0.5 = 0.0506 \\ H \quad H & \end{array}$$

$$\begin{array}{cc} BC \to FC & \\ \downarrow \quad \downarrow & 0.4166 \cdot P(FC|BC) \cdot P(H|FC) = 0.4166 \cdot 0.34 \cdot 0.5 = 0.0708 \\ H \quad H & \end{array}$$

$$\begin{array}{cc} FC \to BC & \\ \downarrow \quad \downarrow & 0.2024 \cdot P(BC|FC) \cdot P(H|BC) = 0.2024 \cdot 0.5 \cdot 0.7 = 0.0708 \\ H \quad H & \end{array}$$

$$\begin{array}{cc} BC \to BC & \\ \downarrow \quad \downarrow & 0.4166 \cdot P(BC|BC) \cdot P(H|BC) = 0.4166 \cdot 0.66 \cdot 0.7 = 0.1925 \\ H \quad H & \end{array}$$

Three Sequence Permutations :

$$\begin{array}{ccc} FC & \to FC & \to FC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.0506 \cdot P(FC|FC) \cdot P(T|FC) = 0.0506 \cdot 0.5 \cdot 0.5 = 0.01265$$

$$\begin{array}{ccc} FC & \to FC & \to BC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.0506 \cdot P(BC|FC) \cdot P(T|BC) = 0.0506 \cdot 0.5 \cdot 0.3 = 0.0076$$

$$\begin{array}{ccc} FC & \to BC & \to FC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.0708 \cdot P(FC|BC) \cdot P(T|FC) = 0.0708 \cdot 0.34 \cdot 0.5 = 0.0120$$

$$\begin{array}{ccc} FC & \to BC & \to BC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.0708 \cdot P(BC|BC) \cdot P(T|BC) = 0.0708 \cdot 0.66 \cdot 0.3 = 0.0140$$

$$\begin{array}{ccc} BC & \to FC & \to FC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.0708 \cdot P(FC|FC) \cdot P(T|FC) = 0.0708 \cdot 0.5 \cdot 0.5 = 0.0177$$

$$\begin{array}{ccc} BC & \to FC & \to BC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.0708 \cdot P(BC|FC) \cdot P(T|BC) = 0.0708 \cdot 0.5 \cdot 0.3 = 0.0106$$

$$\begin{array}{ccc} BC & \to BC & \to FC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.1925 \cdot P(FC|BC) \cdot P(T|FC) = 0.1925 \cdot 0.34 \cdot 0.5 = 0.0327$$

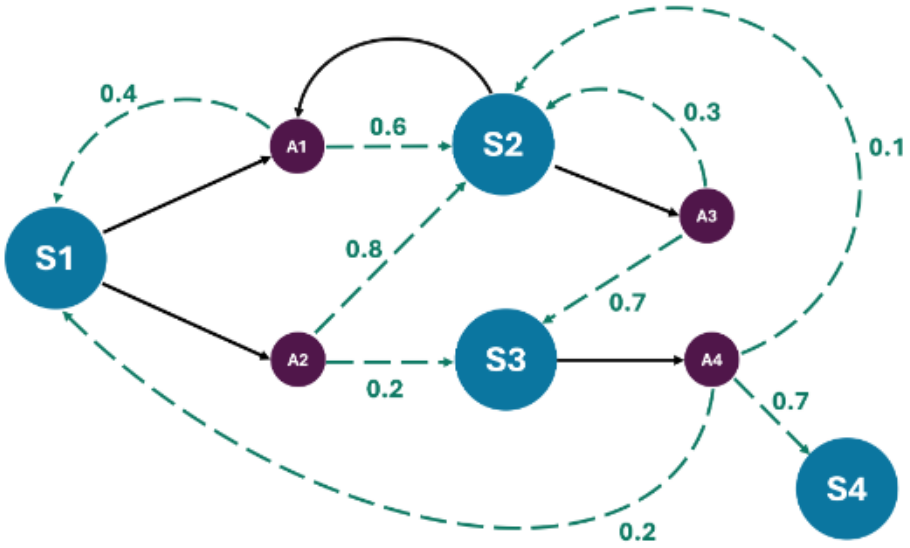
$$\begin{array}{ccc} BC & \to BC & \to BC \\ \downarrow & \downarrow & \downarrow \\ H & H & T \end{array} \quad 0.1925 \cdot P(BC|BC) \cdot P(T|BC) = 0.1925 \cdot 0.66 \cdot 0.3 = 0.0381$$

According to π, it is more likely that the biased coin was flipped in the initial round.

$\therefore$ The most likely sequence of coin types that were flipped to get H, H, T were BC, BC, BC with a probability of 0.0381.

3 Markov Decision Processes and Reinforcement Learning

Consider the following MDP with 4 states, and 4 actions. A dashed line represents a transition from a chosen action to some next state. Transition probabilities are specified along each dashed edge.



Q8 **How many unique policies does this MDP have? Explain your reasoning and list all policies. Use X to indicate no possible action from a state.** **(2)**

S1	S2	S3	S4
A1	A1	A4	X
A1	A3	A4	X
A2	A1	A4	X
A2	A3	A4	X

Table 1: Table showing 4 unique policies

We can create a table to see which combination of policies each state can take and we end up with 4 unique policies.

Q9 **If from any state, all valid actions are equally likely, then what is the total probability of reaching S4 from S1 using paths of at most length 3? List all such paths and compute the total probability. Show your calculations. (An action followed by a transition into a next state counts as a total of one move.)** **(4)**

Path 1 : $S1 \rightarrow (A1\ \ S2) \rightarrow (A3\ \ S3) \rightarrow (A4\ \ S4)$

$= P(A1|S1) \cdot P(S2|A1) \cdot P(A3|S2) \cdot P(S3|A3) \cdot P(A4|S3) \cdot P(S4|A4)$

$= 0.5 \cdot 0.6 \cdot 0.5 \cdot 0.7 \cdot 1.0 \cdot 0.7$

$= 0.0735$

Path 2 : $S1 \rightarrow (A2\ \ S2) \rightarrow (A3\ \ S3) \rightarrow (A4\ \ S4)$

$= P(A2|S1) \cdot P(S2|A2) \cdot P(A3|S2) \cdot P(S3|A3) \cdot P(A4|S3) \cdot P(S4|A4)$

$= 0.5 \cdot 0.8 \cdot 0.5 \cdot 0.7 \cdot 1.0 \cdot 0.7$

$= 0.098$

Path 3 : $S1 \rightarrow (A2\ \ S3) \rightarrow (A4\ \ S4)$

$= P(A2|S1) \cdot P(S3|A2) \cdot P(A4|S3) \cdot P(S4|A4)$

$= 0.5 \cdot 0.2 \cdot 1.0 \cdot 0.7$

$= 0.07$

$\therefore$ There is a $0.0735 + 0.098 + 0.07 = 0.2415$ probability that we can reach S4 from S1 using paths of at most length 3.

Q10 **Given that $R(S1, A1, S2) = 10$, $R(S1, A2, S2) = 10$, $R(S1, A2, S3) = 15$, $R(S3, A4, S4) = 100$, and that rewards for all other transitions are 0, write and expand the optimal value function equation for $V_{opt}(S1)$. Assume that the discount factor is γ, and leave your final answer in terms of $V_{opt}(S2)$ and $V_{opt}(S3)$.** **(4)**

$$V_{opt}(S1) = \max_a \left(\sum_{s'} T(S1, a, s')[R(S1, a, s') + \gamma V_{opt}(s')] \right)$$

For A1:

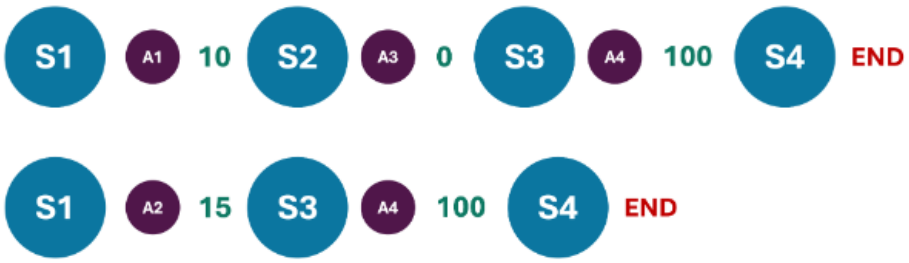
$$\begin{aligned}
&= T(S1, A1, S1)[R(S1, A1, S1) + \gamma V_{opt}(S1)] + T(S1, A1, S2)[R(S1, A1, S2) + \gamma V_{opt}(S2)] \\
&= 0.4[0 + \gamma V_{opt}(S1)] + 0.6[10 + \gamma V_{opt}(S2)] \\
&= (\gamma 0.4 V_{opt}(S1)) + (6 + \gamma 0.6 V_{opt}(S2)) \\
&= 6 + \gamma(0.4 V_{opt}(S1) + 0.6 V_{opt}(S2))
\end{aligned}$$

For A2:

$$\begin{aligned}
&= T(S1, A2, S2)[R(S1, A2, S2) + \gamma V_{opt}(S2)] + T(S1, A2, S3)[R(S1, A2, S3) + \gamma V_{opt}(S3)] \\
&= 0.8[10 + \gamma V_{opt}(S2)] + 0.2[15 + \gamma V_{opt}(S3)] \\
&= (8 + \gamma 0.8 V_{opt}(S2)) + (3 + \gamma 0.2 V_{opt}(S3)) \\
&= 11 + \gamma(0.8 V_{opt}(S2) + 0.2 V_{opt}(S3))
\end{aligned}$$

$$\therefore V_{opt}(S1) = \max_a \Big(6 + \gamma(0.4 V_{opt}(S1) + 0.6 V_{opt}(S2)), 11 + \gamma(0.8 V_{opt}(S2) + 0.2 V_{opt}(S3)) \Big)$$

Q11 **Assume that by simulating this MDP using some exploration policy π, we obtain the two following episodes:**



Use Q-learning updates to calculate the agent's final optimal policy given this data stream, and show all intermediate steps. Assume $\gamma = 1$. For your reference, the Q-learning update equation is given by: **(10)**

$$\eta = \frac{1}{1 + \text{number of updates to } \hat{Q}_{opt}(s, a)}$$

For each observed (s, a, r, s'):

$$Estimate,\ \hat{Q}^{(t)}_{opt}(s, a) = (1 - \eta)\hat{Q}^{(t-1)}_{opt}(s, a) + \eta[R(s, a, s') + \gamma \hat{V}^{(t-1)}_{opt}(s')]$$

$$where\ \hat{V}_{opt}(s') = \max_{a'}\ \hat{Q}_{opt}(s', a')$$

$t = 1 \rightarrow Number\ of\ updates\ for\ (S1, A1) : 0$

$(S1, A2) : 0$

$(S2, A1) : 0$

$(S2, A3) : 0$

$(S3, A4) : 0$

$$\eta(S1, A1) = \frac{1}{1+0} = 1$$

$$\hat{Q}^{(1)}_{opt}(S1, A1) = 0 + 1\left[10 + \max\left(\hat{Q}^{(0)}_{opt}(S1, A1), \hat{Q}^{(0)}_{opt}(S1, A2), \hat{Q}^{(0)}_{opt}(S2, A1), \hat{Q}^{(0)}_{opt}(S2, A3)\right)\right]$$
$$= 0 + 1[10 + 0]$$
$$= 10$$

	t=0	t=1	t=2	t=3	t=4	t=5
$\hat{Q}^{(t)}_{opt}(S1, A1)$	0	10	-	-	-	-
$\hat{Q}^{(t)}_{opt}(S1, A2)$	0	0	-	-	-	-
$\hat{Q}^{(t)}_{opt}(S2, A1)$	0	0	-	-	-	-
$\hat{Q}^{(t)}_{opt}(S2, A3)$	0	0	-	-	-	-
$\hat{Q}^{(t)}_{opt}(S3, A4)$	0	0	-	-	-	-

$t = 2 \rightarrow Number\ of\ updates\ for\ (S1, A1) : 1$

$(S1, A2) : 0$

$(S2, A1) : 0$

$(S2, A3) : 0$

$(S3, A4) : 0$

$$\eta(S2, A3) = \frac{1}{1+0} = 1$$

$$\hat{Q}^{(2)}_{opt}(S2, A3) = 0 + 1\left[0 + \max\left(\hat{Q}^{(1)}_{opt}(S2, A1), \hat{Q}^{(1)}_{opt}(S2, A3), \hat{Q}^{(1)}_{opt}(S3, A4)\right)\right]$$
$$= 0 + 1[0 + 0]$$
$$= 0$$

	t=0	t=1	t=2	t=3	t=4	t=5
$\hat{Q}^{(t)}_{opt}(S1, A1)$	0	10	10	-	-	-
$\hat{Q}^{(t)}_{opt}(S1, A2)$	0	0	0	-	-	-
$\hat{Q}^{(t)}_{opt}(S2, A1)$	0	0	0	-	-	-
$\hat{Q}^{(t)}_{opt}(S2, A3)$	0	0	0	-	-	-
$\hat{Q}^{(t)}_{opt}(S3, A4)$	0	0	0	-	-	-

$t = 3 \rightarrow Number\ of\ updates\ for\ (S1, A1) : 1$

$(S1, A2) : 0$

$(S2, A1) : 0$

$(S2, A3) : 1$

$(S3, A4) : 0$

$$\eta(S3, A4) = \frac{1}{1+0} = 1$$

$$\hat{Q}^{(3)}_{opt}(S3, A4) = 0 + 1\left[100 + \max\left(\hat{Q}^{(2)}_{opt}(S1, A1), \hat{Q}^{(2)}_{opt}(S1, A2), \hat{Q}^{(2)}_{opt}(S2, A1), \hat{Q}^{(2)}_{opt}(S2, A3)\right)\right]$$
$$= 0 + 1[100 + 10]$$
$$= 110$$

	t=0	t=1	t=2	t=3	t=4	t=5
$\hat{Q}^{(t)}_{opt}(S1, A1)$	0	10	10	10	-	-
$\hat{Q}^{(t)}_{opt}(S1, A2)$	0	0	0	0	-	-
$\hat{Q}^{(t)}_{opt}(S2, A1)$	0	0	0	0	-	-
$\hat{Q}^{(t)}_{opt}(S2, A3)$	0	0	0	0	-	-
$\hat{Q}^{(t)}_{opt}(S3, A4)$	0	0	0	110	-	-

$t = 4 \rightarrow Number\ of\ updates\ for\ (S1, A1) : 1$

$(S1, A2) : 0$

$(S2, A1) : 0$

$(S2, A3) : 1$

$(S3, A4) : 1$

$$\eta(S1, A2) = \frac{1}{1+0} = 1$$

$$\hat{Q}^{(4)}_{opt}(S1, A2) = 0 + 1\left[15 + \max\left(\hat{Q}^{(3)}_{opt}(S2, A1), \hat{Q}^{(3)}_{opt}(S2, A3), \hat{Q}^{(3)}_{opt}(S3, A4)\right)\right]$$
$$= 0 + 1\,[15 + 110]$$
$$= 125$$

	t=0	t=1	t=2	t=3	t=4	t=5
$\hat{Q}^{(t)}_{opt}(S1, A1)$	0	10	10	10	10	-
$\hat{Q}^{(t)}_{opt}(S1, A2)$	0	0	0	0	125	-
$\hat{Q}^{(t)}_{opt}(S2, A1)$	0	0	0	0	0	-
$\hat{Q}^{(t)}_{opt}(S2, A3)$	0	0	0	0	0	-
$\hat{Q}^{(t)}_{opt}(S3, A4)$	0	0	0	110	110	-

$t = 5 \rightarrow Number\ of\ updates\ for\ (S1, A1) : 1$

$(S1, A2) : 1$

$(S2, A1) : 0$

$(S2, A3) : 1$

$(S3, A4) : 1$

$$\eta(S3, A4) = \frac{1}{1+1} = \frac{1}{2}$$

$$\begin{aligned}\hat{Q}^{(5)}_{opt}(S3, A4) &= \frac{1}{2}(110) + \frac{1}{2}\left[100 + \max\Big(\hat{Q}^{(4)}_{opt}(S1, A1),\right. \\ &\quad \left.\hat{Q}^{(4)}_{opt}(S1, A2), \hat{Q}^{(4)}_{opt}(S2, A1), \hat{Q}^{(4)}_{opt}(S2, A3)\Big)\right] \\ &= 55 + \frac{1}{2}\left[100 + 125\right] \\ &= 167.5\end{aligned}$$

	t=0	t=1	t=2	t=3	t=4	t=5
$\hat{Q}^{(t)}_{opt}(S1, A1)$	0	10	10	10	10	10
$\hat{Q}^{(t)}_{opt}(S1, A2)$	0	0	0	0	125	125
$\hat{Q}^{(t)}_{opt}(S2, A1)$	0	0	0	0	0	0
$\hat{Q}^{(t)}_{opt}(S2, A3)$	0	0	0	0	0	0
$\hat{Q}^{(t)}_{opt}(S3, A4)$	0	0	0	110	110	167.5

$\therefore$ the agent's final optimal policy given this data stream would be:

$$\pi_{opt} = \Big\{S1 : A2,\ S2 : A1,\ S3 : A4,\ S4 : X\Big\}$$

4 Academic Integrity

Q12 **Review, and copy/paste the following academic integrity acknowledgement in your final submission as the answer to Q12.**

I have read and understood the academic integrity policy as outlined in the course syllabus for CS4100. By pasting this acknowledgement in my submission, I declare that all work presented here is my own, and any conceptual discussions I may have had with classmates have been fully disclosed. I declare that generative AI was not used to answer any questions in this assignment. Any use of generative AI to improve writing clarity alone is accompanied by an appendix with my original, unedited answers.

References

[1] R. Venkatesaramani, "Personal website." https://rajagopalvenkat.com/. Accessed: 2024-07-23.