Machine Learning

Assignment 11.4

Submitted By: Ranji Raj
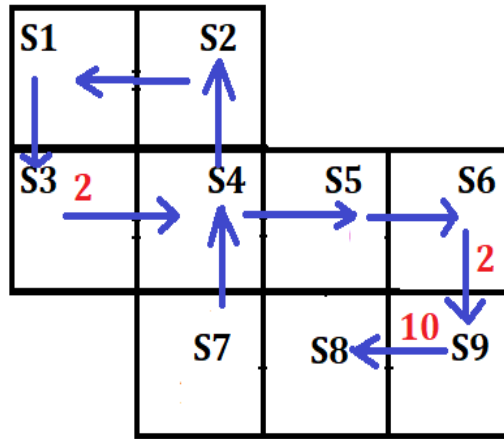
January 23, 2021

Q-Learning

Figure 1: Episode

- Path of the agent: $s_7 - s_4 - s_2 - s_1 - s_3 - s_4 - s_5 - s_6 - s_9 - s_8$

- Update entry table for $\hat{Q}(s, a) \leftarrow r + \gamma \arg\max_{a'} \ \hat{Q}(s', a')$

Interpretation

Not sufficient to repeat this specific episode, reason being some state transitions are never tested at all. e.g. a downward transition from s_2. So the $\hat{Q}$ for that transition will always be 0 and agent might miss the optimal path.

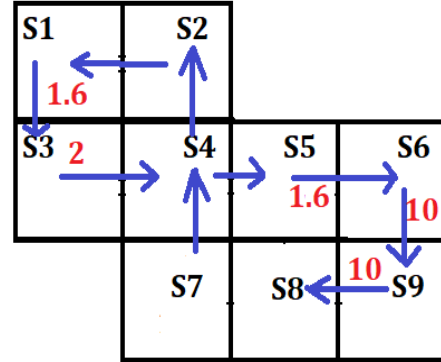| S_{start}	S_{stop}	**calculation** $r + \gamma V^*(S_{stop})$	$\hat{Q}(s', a')$
S_7	S_4	$0 + 0.8(0)$	0.0
S_4	S_2	$0 + 0.8(0)$	0.0
S_2	S_1	$0 + 0.8(0)$	0.0
S_1	S_3	$0 + 0.8(2)$	1.6
S_3	S_4	$2 + 0.8(0)$	2.0
S_4	S_5	$0 + 0.8(0)$	0.0
S_5	S_6	$0 + 0.8(2)$	1.6
S_6	S_9	$2 + 0.8(10)$	10.0
S_9	S_8	$10 + 0.8(0)$	10.0

Table 1: $\hat{Q}$-table-1

Table 2: Episode-1

Table 3: Iteration-1

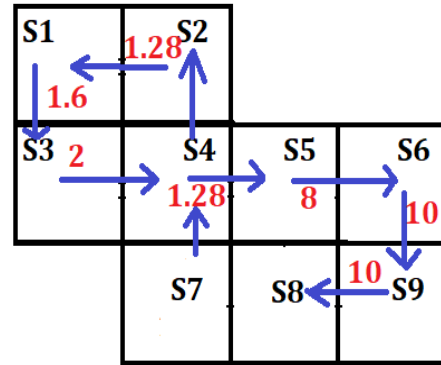| S_{start}	S_{stop}	**calculation** $r + \gamma V^*(S_{stop})$	$\hat{Q}(s', a')$
S_7	S_4	$0 + 0.8(0)$	0.0
S_4	S_2	$0 + 0.8(0)$	0.0
S_2	S_1	$0 + 0.8(1.6)$	1.28
S_1	S_3	$0 + 0.8(2)$	1.6
S_3	S_4	$2 + 0.8(0)$	2.0
S_4	S_5	$0 + 0.8(1.6)$	1.28
S_5	S_6	$0 + 0.8(10)$	8.00
S_6	S_9	$2 + 0.8(10)$	10.0
S_9	S_8	$10 + 0.8(0)$	10.0

Table 4: $\hat{Q}$-table-2

Table 5: Episode-2

Table 6: Iteration-2

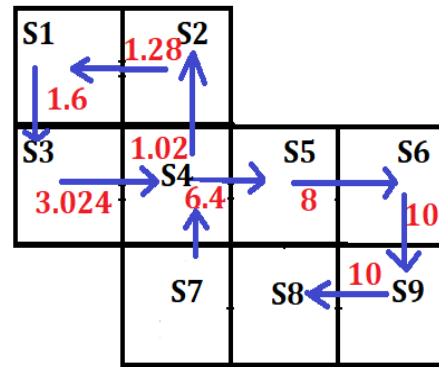| S_{start}	S_{stop}	calculation $r + \gamma V^*(S_{stop})$	$\hat{Q}(s', a')$
S_7	S_4	$0 + 0.8(0)$	0.0
S_4	S_2	$0 + 0.8(1.28)$	1.024
S_2	S_1	$0 + 0.8(1.6)$	1.28
S_1	S_3	$0 + 0.8(2)$	1.6
S_3	S_4	$2 + 0.8(1.28)$	3.024
S_4	S_5	$0 + 0.8(8)$	6.4
S_5	S_6	$0 + 0.8(10)$	8.00
S_6	S_9	$2 + 0.8(10)$	10.0
S_9	S_8	$10 + 0.8(0)$	10.0

Table 7: $\hat{Q}$-table-3

Table 8: Episode-3

Table 9: Iteration-3