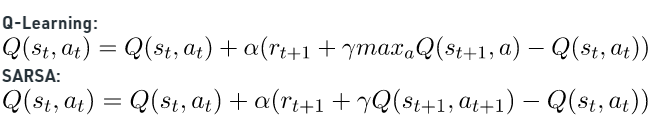
1.

2.



Q-Learning:

Q(State1, Up) = 2 + 0.9 \* (-1 + 0.8 \* 4 – 2) = 2.18

Q(State2, Right) = 4 + 0.9 \* (-1 + 0.8 \* 6 – 4) = 3.82

Q(State3, Right) = 6 + 0.9 \* (-1 + 0.8 \* 10 – 6) = 6.9

Q(State4, Down) = -40 + 0.9 \* (-100 + 0.8 \* 0 – (-40)) = -94

SARSA:

Q(State1, Up) = 2 + 0.9 \* (-1 + 0.8 \* 4 – 2) = 2.18

Q(State2, Right) = 4 + 0.9 \* (-1 + 0.8 \* 6 – 4) = 3.82

Q(State3, Right) = 6 + 0.9 \* (-1 + 0.8 \* (-40) – 6) = -29.1

Q(State4, Down) = -40 + 0.9 \* (-100 + 0.8 \* 0 – (-40)) = -94

3.





first -visit:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1:0 | 2:0 | 3:0 | 0 | 0 |
| 0 | 7:0 | 8:0 | 0 | 0 |
| 0 | 12:0 |  | 0 | 0 |
| 0 | 16:0 |  | 17:5 | 18:5 |
| 0 | 20:0 | 21:2.5 | 22:2.5 | 23: |

every-visit:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1:0 | 2:0 | 3:0 | 0 | 0 |
| 0 | 7:0 | 8:0 | 0 | 0 |
| 0 | 12:0 |  | 0 | 0 |
| 0 | 16:0 |  | 17:5 | 18:5 |
| 0 | 20:0 | 21:3.75 | 22:3.75 | 23:0 |