Value iteration and policy iteration

Tuesday, 4 September 2018 9:34 AM

$$V(s) = \max a \in A(s) \sum_{s' \in S} P_a(s'|s)[r(s,a,s') + \gamma V(s')]$$

Value iteration:

```
1) Set V_0 to arbitrary value for each s in S (choose 0 as the value)
2) While diff is >= epsilon
    a. For each s in S do
        i. V_t+1(s) := $\max a \in A(s) \sum_{s' \in S} P_a(s'|s)[r(s,a,s') + \gamma V\_t(s')]$
3) Select policy
```

0.8*(0 + 0.9*0) = 0

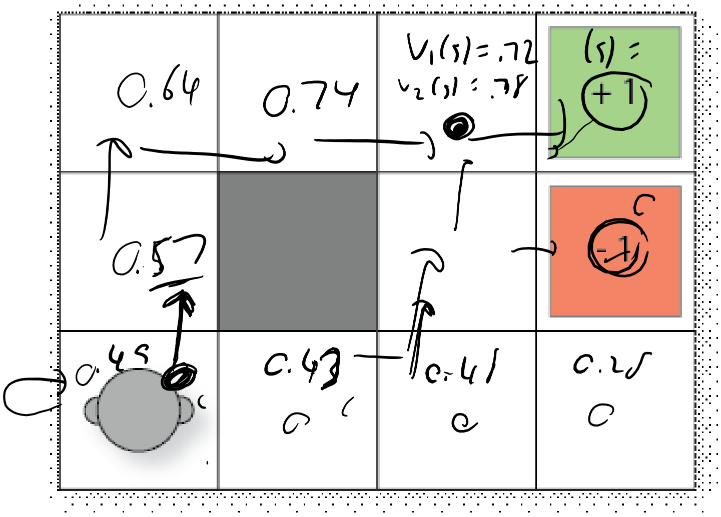
0.1*(0 + 0.9*0) = 0

0.1*(0 + 0.9*1) = 0.09

0.06

$0.72 + 0 + 0.1(0 + 0.9 \times 0.72)$

$0.8(0 + 0.9 \times 0.72)$

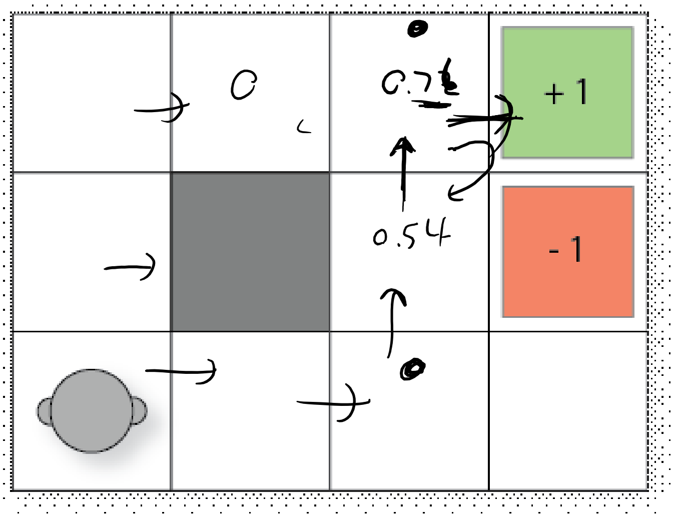


The two labelled cells give a reward: 1 and -1 respectively. (Actually, we will assume V(s)=1 or -1)

But! Things can go wrong:

- If the agent tries to move north, 80% of the time, this works as planned (provided the wall is not in the way)
- 10% of the time, trying to move north takes the agent west (provided the wall is not in the way);
- 10% of the time, trying to move north takes the agent east (provided the wall is not in the way)
- If the wall is in the way of the cell that would have been taken, the agent stays put.
- Similar for all other directions

Policy iteration:



$V^{\pi}(s) =$

.