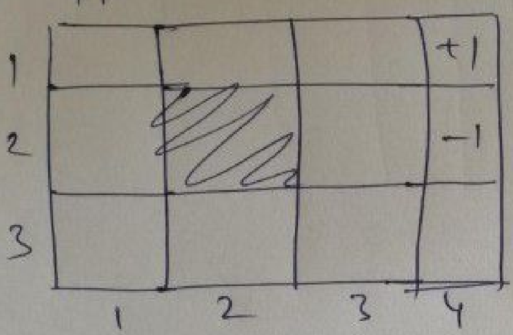
Grid world Example

Demo of value iteration algorithm.

γ (discount factor) = 0.9   living Reward = 0
ie. $R(S) = 0$

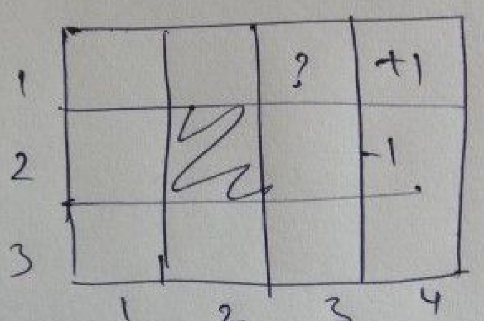
V_1



0.8 prob. of heading in optimal direction

0.1 + 0.1 prob of heading in either left / Right.

V_2



$V(3,1) = ??$

optimal direction action ⇒ to take right $V(4,1)$

$V(3,1) \longrightarrow V(4,1)$

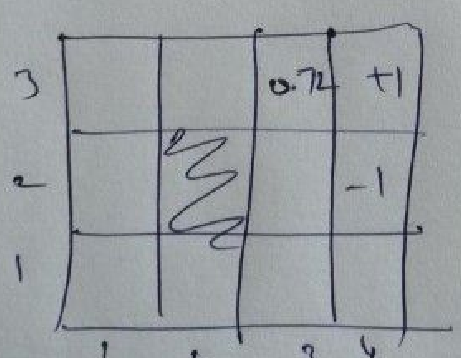
$V(3,1) = R(3,1) + \gamma \sum P \, V(s')$

0.8 $V(4,1)$
0.1   hitting against the wall and coming back)
0.1 $V(3,2)$

$= 0 + 0.9 \left[0.8 \times 1 + 0.1 \times 0 + 0.1 \times 0 \right]$

$= 0.72$

next iteration



now cal $V(3,3)$, $V(3,2)$, $V(2,3)$

$V(3,3) = R(3,1) + \gamma \sum p V(s')$

$= 0 + \underset{0.9}{\gamma} \left[0.8 \times 1 + \underbrace{0.1 \times 0.72}_{} + 0.1 \times 0 \right]$

Agent hits against the wall and comes back

$= 0.78$

$V(3,2) = R(3,2) + \gamma \sum P\, V(s')$

0.8 → Taking right action going to V(3,3)

0.1 → hitting against the wall

0.1 → going to V(4,2)

$= 0 + 0.9 \left[0.8 \times 0.72 + 0.1 \times 0 + 0.1 \times (-1) \right]$

$= 0.4284 \approx 0.43$

$V(2,3) = R(2,3) + \gamma \sum P\, V(s')$

0.8 → Taking right action going to V(3,3)

0.1 → hitting against the wall

0.1 → hitting against the blocked cell (V(2,2))

$= 0 + 0.9 \left[0.8 \times 0.72 + 0.1 \times 0 + 0.1 \times 0 \right]$

$= 0.5184 \approx 0.52$

0	0.52	0.78	+1
0		0.43	-1
0	0	0	0

⋮

You carry on going through the next set of iterations until value functions for all states in the grid are calculated.