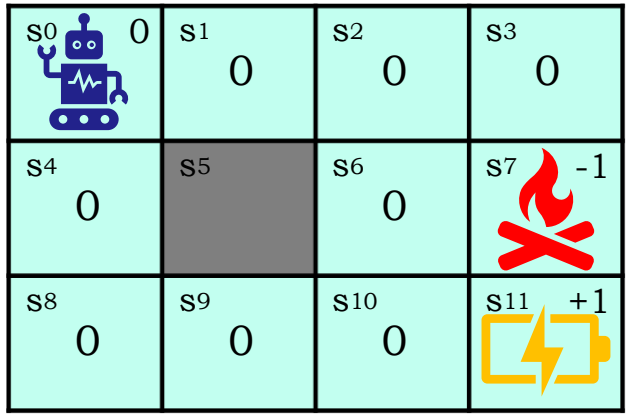
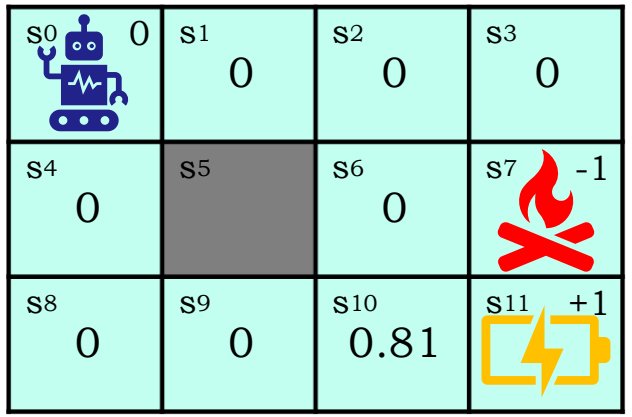
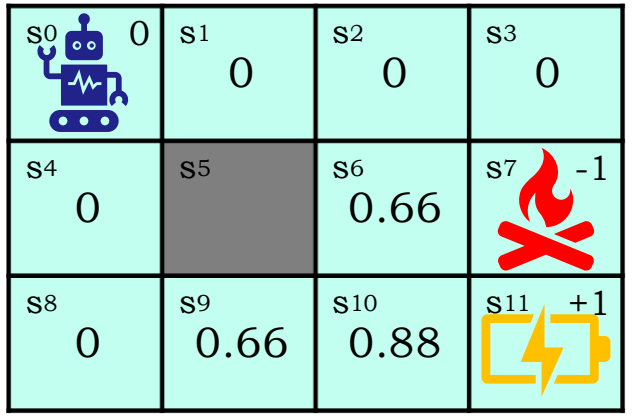
Value Iteration: Dynamic Programming

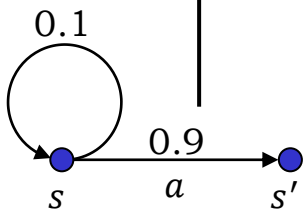
$State, s \in S$
$Action, a \in A(s)$
$Reward, r \in R(s, a, s')$
$Transition\ Probability, P(s, a, s')$
$Optimal\ Policy, \pi^*(s)$
$Return, G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$
$State\ value, V^{\pi}(s)$
$Action\ value, Q^{\pi}(s, a)$
$V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$
$V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s, a)[r + \gamma V^*_k(s')]$
$Q^*(s, a) = \sum_{s'} p(s'|s, a)\left[r + \gamma \max_{a'} Q^*(s', a')\right]$

- Bellman's optimal value function
  - $V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$
- Bellman update
  - $V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s, a)[r + \gamma V^*_k(s')]$
  - $V_2(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0] = 0.81$
  - $V_2(s_0, \dots, s_9) = 0$
  - $V_3(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0.81] = 0.8843$
  - $V_3(s_9) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$
  - $V_3(s_6) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$







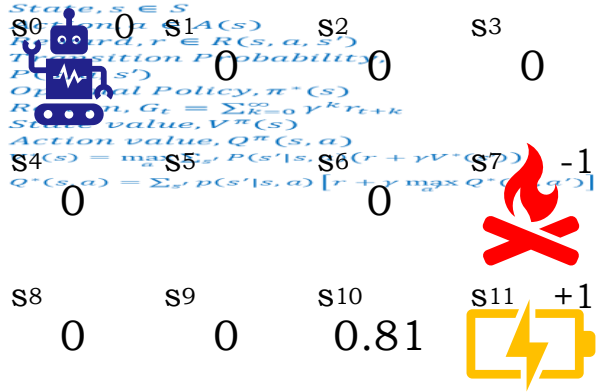


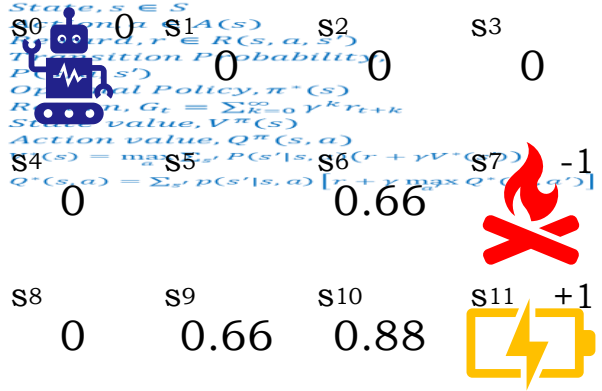
https://www.profkumar.co.in/

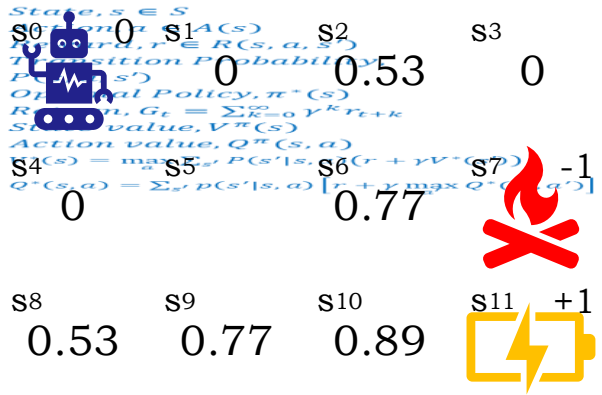
Value Iteration: Dynamic Programming

$State, s \in S$
$Action, a \in A(s)$
$Reward, r \in R(s,a,s')$
$Transition\ Probability, P(s,a,s')$
$Optimal\ Policy, \pi^*(s)$
$Return, G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$
$State\ value, V^{\pi}(s)$
$Action\ value, Q^{\pi}(s,a)$
$V^*(s) = \max_a \sum_{s'} P(s'|s,a)\left(r + \gamma V^*(s')\right)$
$Q^*(s,a) = \sum_{s'} p(s'|s,a)\left[r + \gamma \max_{a'} Q^*(s',a')\right]$

- Bellman's optimal value function
  - $V^*(s) = \max_a \sum_{s'} P(s'|s,a)\left(r + \gamma V^*(s')\right)$
- Bellman update
  - $V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s,a)[r + \gamma V^*_k(s')]$
  - $V_2(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0] = 0.81$
  - $V_2(s_0, \dots, s_9) = 0$
  - $V_3(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0.81] = 0.8843$
  - $V_3(s_9) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$
  - $V_3(s_6) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$







https://www.profkumar.co.in/

Value Iteration: Dynamic Programming

$State, s \in S$
$Action, a \in A(s)$
$Reward, r \in R(s, a, s')$
$Transition\ Probability, P(s, a, s')$
$Optimal\ Policy, \pi^*(s)$
$Return, G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$
$State\ value, V^{\pi}(s)$
$Action\ value, Q^{\pi}(s, a)$
$V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$
$Q^*(s, a) = \sum_{s'} p(s'|s, a)\left[r + \gamma \max_{a'} Q^*(s', a')\right]$

- Bellman's optimal value function
  - $V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$
- Bellman update
  - $V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s, a)[r + \gamma V^*_k(s')]$
  - $V_2(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0] = 0.81$
  - $V_2(s_0, \dots, s_9) = 0$
  - $V_3(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0.81] = 0.8843$
  - $V_3(s_9) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$
  - $V_3(s_6) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$

https://www.profrkumar.co.in/

s0 0	s1 0	s2 0	s3 0
s4 0	s5	s6 0.66	s7 -1
s8 0	s9 0.66	s10 0.88	s11 +1

s0 0	s1 0	s2 0.53	s3 0
s4 0	s5	s6 0.77	s7 -1
s8 0.53	s9 0.77	s10 0.89	s11 +1

s0 0	s1 0.43	s2 0.67	s3 0.43
s4 0.43	s5	s6 0.79	s7 -1
s8 0.67	s9 0.79	s10 0.89	s11 +1

Value Iteration: Dynamic Programming

$State, s \in S$

$Action, a \in A(s)$

$Reward, r \in R(s, a, s')$

$Transition\ Probability, P(s, a, s')$

$Optimal\ Policy, \pi^*(s)$

$Return, G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$

$State\ value, V^{\pi}(s)$

$Action\ value, Q^{\pi}(s, a)$

$V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$

$V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s, a)[r + \gamma V^*_k(s')]$

$Q^*(s, a) = \sum_{s'} p(s'|s, a)\left[r + \gamma \max_{a'} Q^*(s', a')\right]$

- Bellman's optimal value function
  - $V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$
- Bellman update
  - $V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s, a)[r + \gamma V^*_k(s')]$
  - $V_2(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0] = 0.81$
  - $V_2(s_0, \dots, s_9) = 0$
  - $V_3(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0.81] = 0.8843$
  - $V_3(s_9) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$
  - $V_3(s_6) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$

https://www.profrkumar.co.in/

s0 0	s1 0	s2 0.53	s3 0
s4 0	s5	s6 0.77	s7 -1
s8 0.53	s9 0.77	s10 0.89	s11 +1

s0 0	s1 0.43	s2 0.67	s3 0.43
s4 0.43	s5	s6 0.79	s7 -1
s8 0.67	s9 0.79	s10 0.89	s11 +1

s0 0.35	s1 0.58	s2 0.7	s3 0.58
s4 0.58	s5	s6 0.79	s7 -1
s8 0.7	s9 0.79	s10 0.89	s11 +1

Value Iteration: Dynamic Programming

$State, s \in S$

$Action, a \in A(s)$

$Reward, r \in R(s, a, s')$

$Transition\ Probability, P(s, a, s')$

$Optimal\ Policy, \pi^*(s)$

$Return, G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$

$State\ value, V^{\pi}(s)$

$Action\ value, Q^{\pi}(s, a)$

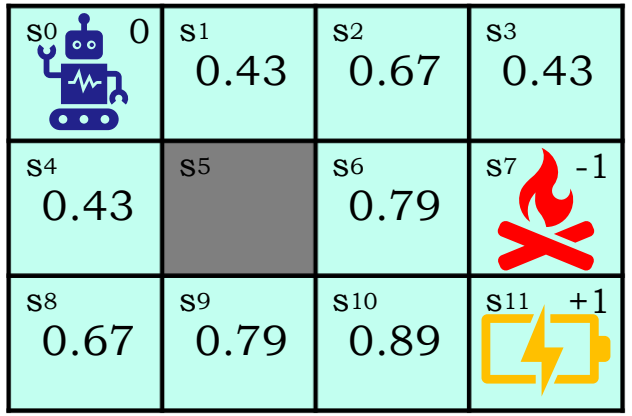
$V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$

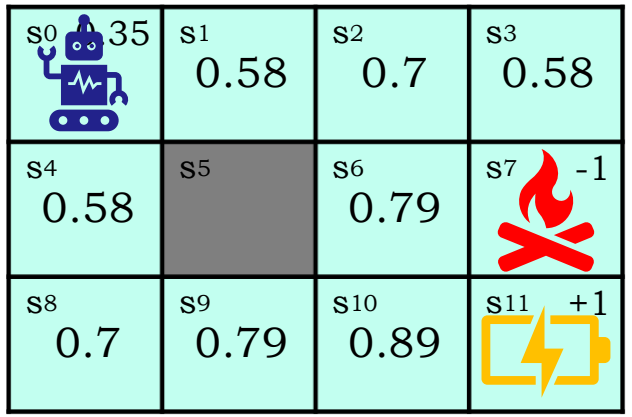
$V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s, a)[r + \gamma V^*_k(s')]$

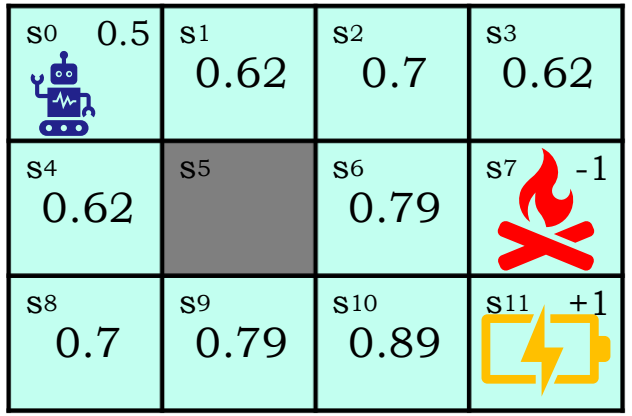
$Q^*(s, a) = \sum_{s'} p(s'|s, a)\left[r + \gamma \max_{a'} Q^*(s', a')\right]$

- Bellman's optimal value function
  - $V^*(s) = \max_a \sum_{s'} P(s'|s, a)\big(r + \gamma V^*(s')\big)$
- Bellman update
  - $V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s, a)[r + \gamma V^*_k(s')]$
  - $V_2(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0] = 0.81$
  - $V_2(s_0, \dots, s_9) = 0$
  - $V_3(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0.81] = 0.8843$
  - $V_3(s_9) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$
  - $V_3(s_6) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$

https://www.profrkumar.co.in/



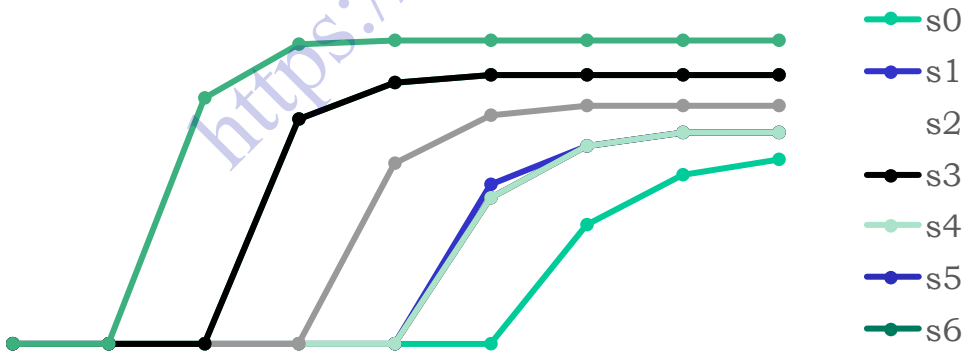




Value Iteration: Dynamic Programming

$State, s \in S$
$Action, a \in A(s)$
$Reward, r \in R(s,a,s')$
$Transition\ Probability, P(s,a,s')$
$Optimal\ Policy, \pi^*(s)$
$Return, G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$
$State\ value, V^{\pi}(s)$
$Action\ value, Q^{\pi}(s,a)$
$V^*(s) = \max_a \sum_{s'} P(s'|s,a)\big(r + \gamma V^*(s')\big)$
$V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s,a)[r + \gamma V^*_k(s')]$
$Q^*(s,a) = \sum_{s'} p(s'|s,a)\left[r + \gamma \max_{a'} Q^*(s',a')\right]$

- Bellman's optimal value function
  - $V^*(s) = \max_a \sum_{s'} P(s'|s,a)\big(r + \gamma V^*(s')\big)$
- Bellman update
  - $V^*_{k+1}(s) = \max_a \sum_{s'} P(s'|s,a)[r + \gamma V^*_k(s')]$
  - $V_2(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0] = 0.81$
  - $V_2(s_0, \dots, s_9) = 0$
  - $V_3(s_{10}) = 0.9[0 + 0.9\times1] + 0.1[0 + 0.9\times0.81] = 0.8843$
  - $V_3(s_9) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$
  - $V_3(s_6) = 0.9[0 + 0.9\times0.81] + 0.1[0 + 0.9\times0] = 0.6561$



s0 0.35	s1 0.58	s2 0.7	s3 0.58
s4 0.58	s5	s6 0.79	s7 -1
s8 0.7	s9 0.79	s10 0.89	s11 +1

s0 0.5	s1 0.62	s2 0.7	s3 0.62
s4 0.62	s5	s6 0.79	s7 -1
s8 0.7	s9 0.79	s10 0.89	s11 +1

s0 0.54	s1 0.62	s2 0.7	s3 0.62
s4 0.62	s5	s6 0.79	s7 -1
s8 0.7	s9 0.79	s10 0.89	s11 +1

https://www.profrkumar.co.in/