Solve MRP

MRP

① S ② $P(s, s')$, ③ R . ④ γ .

⑤ $T = \inf \{t \geq 0 , S_t \in \partial S\}$

Episode

Terminal Reward

$$\omega = \{S_0 , R_1 , S_1 , \cdots R_T , S_T , R_{T+1}\}$$

Gain

$$G_t = R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-t} R_{T+1}$$

Value

$$V(s) = \mathbb{E}[G_t \mid S_t = s]$$

Bellman

$$\begin{cases} v(s) = \mathbb{E}[R_{t+1} + \gamma v(S_{t+1}) \mid S_t = s] & \forall s \notin \partial S \\ v(s) = R(s) & \forall s \in \partial S \end{cases}$$

To solve Bellman

1. linear Algebra
2. DP
3. MC
4. TD

$$v(s) = R(s) + \gamma \sum_{s'} P(s, s') \, v(s')$$

$$\begin{cases} v(s) = \mathbb{E}[R_{t+1} + \gamma v(S_{t+1}) \mid S_t = s]. & \forall s \notin \partial S \\ v(s) = R(s) & \forall s \in \partial S \end{cases}$$

① Lin Alg.

Denote $\{v(s) : s \in S\}$ be a column vector

$\{R(s) : s \in S\}$ — — — — —

write Bellman

$$v = R + \gamma P v$$

$$v = (I - \gamma P)^{-1} R$$

② DP

$$v^{(n+1)} \Longleftarrow R + \gamma P v^{(n)}$$

③ MC I (first visit)

To compute $V(s)$, use

$$V(s) = \mathbb{E}[G_0 \mid S_0 = s]$$

$$\approx \frac{1}{n} \sum_{i=1}^{n} G_0^{(i)}$$

Algo

Tot = 0

for i = 1 ⋯ n do

Generate $w_i = \{ \underset{= s}{S_0}, R_1, S_1, \cdots R_T, S_T, R_{T+1} \}$

Compute $G \leftarrow R_1 + \gamma R_2 + \cdots + \gamma^T R_{T+1}$

Tot ← Tot + G

return Tot/n

③ MC2 – (every visit)

Algo (To compuse $V(s)$)

$Tot = 0$

For $i = 1 \cdots n$ do

$$\omega(i) = \{S_0, R_1, S_1 \cdots R_T, S_T, R_{T+1}\}$$

$$t^i(0) = \inf\{t \geq 0, S_t = s\}$$

$$t^i(j+1) = \inf\{t \geq t^i(j), S_t = s\} \wedge T$$

$$M^i = \#\{0 \leq t < T : S_t = s\}$$

$$V(s) = \frac{1}{\sum_{i=1}^{n} M_i} \sum_{i=1}^{n} \sum_{j=0}^{M^i} G\left(t^i(j)\right)$$

Fact

Let $(x_1, x_2 \cdots)$ be a seq.

$$\mu_k = \frac{1}{k}\sum_{i=1}^{k} x_i$$

Then $\mu_k = \mu_{k-1} + \frac{1}{k}(x_k - \mu_{k-1})$.

Pf (skip)

Algo $\mu = 0$.

for $k = 1, 2 \cdots n$ do

Generate x_k

$\mu \leftarrow \mu + \frac{1}{k}(x_k - \mu)$

α (learning rate)

Return μ.

MC2 algo : To compute $\{v(s) : s \in S\}$ for MRP

Init $v(s) = 0 \quad \forall s \in S$

for $i = 1 \cdots n$ do

Generate $\omega = \{S_0, R_1, S_1, \cdots R_T, S_T, R_{T+1}\}$

$G \leftarrow R_{T+1}$

for $t = T-1 \cdots 0$ do

$G \leftarrow R_{t+1} + \gamma G$.

$v(S_t) \leftarrow v(S_t) + \alpha(G - v(S_t))$

Return v.

"MC2"

$$V(S_t) \leftarrow V(S_t) + \alpha\left(G_t - V(S_t)\right)$$

"TD(0)"

δ_t : TD error

$$V(S_t) \leftarrow V(S_t) + \alpha\left(\overbrace{\underbrace{R_{t+1} + \gamma V(S_{t+1})}_{\text{TD-target}} - V(S_t)}^{\delta_t}\right)$$

pros

No need for a complete episode for update the value

Algo Compute v by TD(0)

Init $v \equiv 0$

for $i = 1 \cdots n$ do

Generate s,

while $s \notin \partial S$ do

Generate $s' \sim P(s, s')$

$$v(s) \leftarrow v(s) + \alpha\left(R(s) + \gamma v(s') - v(s)\right)$$

$s \leftarrow s'$

Return v

Rk TD(n) is generalization of TD(0)