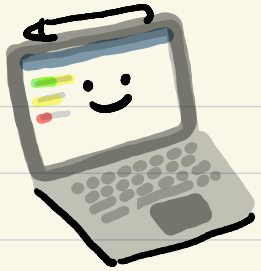

CS234 - Winter 2023

Assignment 1

1) Reward Choices

a) $\hat{r}(s, a) = r(s, a) + c$

New Optimal Value function

$$V'(s) = \max_a E_{\pi_{r'}}\left[\sum_{k=0}^{\infty} \gamma^k \hat{R}_{t+k+1} \mid S_t = s, A_t = a\right]$$

$$= \max_a E_{\pi^*}\left[\sum_{k=0}^{\infty} \gamma^k (R_{t+k+1} + c) \mid S_t = s, A_t = a\right]$$

$$= V(s) + \frac{c}{1-\gamma} \qquad \left(c \sum_{k=0}^{\infty} \gamma^k\right)$$

$$\Rightarrow V'(s) - t = V(s) \quad \text{or} \quad \boxed{V'(s) = V(s) + \frac{c}{1-\gamma}}$$

(+) The optimal policy does not change.

Because t action, the value of state has been increased the same amount $\Rightarrow$ Optimal point doesn't change!

$$\pi'(s) = \underset{a}{\operatorname{argmax}} \sum_{s'} P(s'|s,a)\left(\hat{r}(s,a) + \gamma V'(s')\right)$$

$$= \underset{a}{\operatorname{argmax}} \sum_{s'} P(s'|s,a)\left(r(s,a) + c + \gamma V(s') + \frac{\gamma c}{1-\gamma}\right)$$

$$= \pi(s) \text{ (remove the constant)}$$

b) $\hat{r}(s,a) = c \times r(s,a)$

$$V'(s) = \max_a E_{\pi_a}\left[\sum_{k=0}^{\infty} \gamma^k \hat{R}_{t+k+1} \mid S_t = s, A_t = a\right]$$

$$= \max_a E_{\pi^*}\left[\sum_{k=0}^{\infty} c \times \gamma^k R_{t+k+1} \mid S_1 = s, A_t = a\right]$$

$$= c \times V(s)$$

① $c > 0 \Rightarrow \pi' = \pi^*$

② $c < 0 \Rightarrow \pi' = \pi_1^*$ ($V(s, \pi_1^*)$ has min value)

③ $c = 0 \Rightarrow V'(s) = 0 \ \forall a \Rightarrow$ All policies are optimal.

c) If MDP is finite, does it affect the answer from part a ?

Because MDP is finite, we can ignore γ

$$V'(s) = \max_a \mathbb{E}_{\pi_\gamma}\left[\sum_{k=0}^{M_s} \hat{R}_{t+k+1} \mid S_t = s, A_t = a\right]$$

$$= \max_a \mathbb{E}_{\pi^*}\left[\sum_{k=0}^{M_s} R_{t+k+1} + c \mid S_t = s, A_t = a\right]$$

$$= c \cdot M_s + V(s)$$

M_s : number of steps from state s to the end state

If take γ into account

$$\Rightarrow V'(s) = V(s) + c \cdot \sum_{k=0}^{M_s} \gamma^k$$

Note : M_s is not fixed if the process is stochastic

So the optimal policy will change !

2) Bellman Residuals and performance bounds

B : Bellman backup with fixed point V^* (the optimal value function)

$$(BV)(s) = \max_a \left[r(s,a) + \gamma \sum_{s' \in S} p(s'|s,a) V(s') \right]$$

B^π : Bellman backup with fixed point V^π : (value function follows policy π)

$$(B^\pi V)(s) = \mathop{E}_{a\sim\pi}[r(s,a) + \gamma \sum_{s'\in S} P(s'|s,a)\, V(s')]$$

We did show that $\|BV - BV'\| \leq \gamma \|V - V'\|$

Similarly, we can show $\|B^\pi V - B^\pi V'\| \leq \gamma \|V - V'\|$

a) Prove fixed point of B^π is unique

Assume $m \neq n$ satisfies $B^\pi m = m$, $B^\pi n = n$

$\Rightarrow \|B^\pi m - B n\| = \|m - n\|$

But B^π is a contraction $\Rightarrow \|B^\pi m - B^\pi n\| \leq \gamma \|m - n\| < \|m - n\|$

(contradiction)

-1 Point

b) $(BV)(s) \geq (B^{\pi}V)(s) \quad \forall s \in S, V \in R^{|S|}$

c) When $V = v^{*}$ <optimal value functions> then

$\|BV - V\| = 0$

d)

(+) Proof of $\|V - v^{\pi}\| \leq \frac{\|V - B^{\pi}V\|}{1-\gamma} \quad \forall V$

we know that v^{π} is the fixed point of the contraction B^{π}

$\Rightarrow \lim_{n \to \infty} (B^{\pi})^{n} V = v^{\pi}$

Lemma: $\|(B^\pi)^n V - (B^\pi)^{n-1} V\| \leq \gamma^{n-1} \|B^\pi V - V\|$

Because B^π is a contraction

$\Rightarrow \|(B^\pi)^n V - (B^\pi)^{n-1} V\| \leq \gamma \|(B^\pi)^{n-1} V - (B^\pi)^{n-2} V\|$

$\cdots$

$\leq \gamma^{n-1} \|B^\pi V - V\|$

Back to the main:

$$\|V - V^\pi\| = \lim_{n \to \infty} \|(B^\pi)^n V - V\|$$

$$\leq \lim_{n \to \infty} \|(B^\pi)^n V - (B^\pi)^{n-1} V\| + \|(B^\pi)^{n-1} V - (B^\pi)^{n-2} V\| + \cdots + \|B^\pi V - V\|$$

$$\leq \lim_{n \to \infty} \left(\gamma^{n-1} \|B^\pi V - V\| + \gamma^{n-2} \|B^\pi V - V\| + \cdots + \|B^\pi V - V\| \right)$$

< Apply the lemma >

$$= \|B^{\pi}V - V\| \cdot \lim_{n\to\infty} \sum_{k=0}^{n-1} \gamma^{k}$$

$$= \|V - B^{\pi}V\| \cdot \frac{1}{1-\gamma} \qquad (Q.E.D)$$

Similarly, Because V^* is the fixed Point of contraction B

$$\Rightarrow \|V - BV^*\| \leq \frac{\|V - BV\|}{1-\gamma}$$

e) Proof : $V^{\pi}(s) \geq V^*(s) - \frac{2\varepsilon}{1-\gamma}$ with $\varepsilon = \|BV - V\|$

$$\Leftrightarrow \frac{2\varepsilon}{1-\gamma} \geq V^*(s) - V^{\pi}(s)$$

$$V^*(s) - V^{\pi}(s) = (V^*(s) - V(s)) + (V(s) - V^{\pi}(s))$$

$$\leq \|V^*(s) - V(s)\| + \|V^{\pi}(s) - V(s)\|$$

$$\leq \frac{\|BV - V\|}{1-\gamma} + \frac{\|B^{\pi}V - V\|}{1-\gamma}$$

$$\leq \frac{\|BV - V\|}{1-\gamma} + \frac{\|BV - V\|}{1-\gamma}$$

(Because π is greedy policy)

$B^{\pi}V$ converges faster than BV

Note : if T is the fixed point of contraction B

$\Rightarrow \forall V$ then $\{V, BV, B^2V, \ldots\}$ converges to T

- π^* is the optimal policy with Bellman operator B
  π is an abitrary policy with Bellman operator B^π

=> Convergent speed of $\{V, BV, B^2V, \dots\} > \dots$ of $\{V, B^\pi V, B^{\pi 2} V, \dots\}$

f) Prove $V^\pi(s) \geq v^*(s) - \frac{\varepsilon}{1-\gamma}$ if $V^* \leq V$

we have $V^*(s) - V^\pi(s) \leq V(s) - V^\pi(s)$

$\leq \|V - V^\pi\| \leq \frac{\varepsilon}{1-\gamma}$ (Q.E.D)

g) Show that if $BV \leq V$ then $V^* \leq V$
(Từ đây suy ra tc BV hội tụ)

với mọi không thì V^* (optimal) $\leq V$ (với k)

We prove $B^k V \leq V \quad \forall k \in \mathbb{N}^*$

$$B^k(V)(s) = \max_a \left[r(s,a) + \gamma \sum_{s' \in S} P(s'|s,a) (B^{k-1} V)(s') \right]$$

$$\leq \max_a \left[r(s,a) + \gamma \sum_{s' \in S} P(s'|s,a) V(s') \right]$$

[because $B^{k-1} V \leq V$]

$$= B V \leq V$$

$$\Rightarrow V \geq \lim_{k \to \infty} B^k V = V^*$$

Because V^* is the fixed point of the contraction B.

$$\Rightarrow V \geq V^* \quad (\text{Q.E.D})$$