CSE 250A HW9

Jiping Lin A15058075

1. We have $\log\gamma \le \gamma - 1 \Rightarrow e^{\log\gamma} \le e^{\gamma-1}$

$\Rightarrow \gamma \le e^{\gamma-1}$

then $\sum_{n\ge t} \gamma^n r_n \le \sum_{n\ge t} \gamma^n$ 「Since $0 \le r_n \le 1, \forall n$」

$= \frac{\gamma^t}{1-\gamma}$ 「Geo. Series」

$\le \frac{e^{t(\gamma-1)}}{1-\gamma}$

$= h e^{t(\gamma-1)}$ 「$h = \frac{1}{1-\gamma}$」

$= h e^{-t/h}$

BE: $V^\pi(s) = R(s) + \gamma \sum_{s'} P(s'|s,\pi(s)) \cdot V^\pi(s')$

2.(a) $V^\pi(1) = R(1) + \frac{2}{3}[P(1|1,\uparrow)V^\pi(1) + P(2|1,\uparrow)V^\pi(2) + P(3|1,\uparrow)V^\pi(3)]$

$= -15 + \frac{2}{3}[\frac{3}{4}V^\pi(1) + \frac{1}{4}V^\pi(2)]$

$V^\pi(2) = R(2) + \frac{2}{3}[P(1|2,\uparrow)V^\pi(1) + P(2|2,\uparrow)V^\pi(2) + P(3|2,\uparrow)V^\pi(3)]$

$= 30 + \frac{2}{3}[\frac{1}{2}V^\pi(1) + \frac{1}{2}V^\pi(2)]$

$V^\pi(3) = R(3) + \frac{2}{3}[P(1|3,\downarrow)V^\pi(1) + P(2|3,\downarrow)V^\pi(2) + P(3|3,\downarrow)V^\pi(3)]$

$= -25 + \frac{2}{3}[\frac{1}{4}V^\pi(2) + \frac{3}{4}V^\pi(3)]$

First 2 eq. gives

$V^\pi(1) = -15 + \frac{1}{2}V^\pi(1) + \frac{1}{6}V^\pi(2)$

$V^\pi(2) = 30 + \frac{1}{3}V^\pi(1) + \frac{1}{3}V^\pi(2)$

$\Rightarrow V^\pi(1) = -18,\ V^\pi(2) = 36$. $\Rightarrow V^\pi(3) = -25 + \frac{1}{6}V^\pi(2) + \frac{1}{2}V^\pi(3)$

$\Rightarrow V^\pi(3) = -19$

-18
36
-19

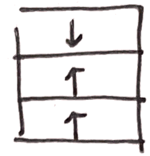
(b) $Q^\pi(s,a)$ are terms in [] in pt. (a),

$\pi'(1) = \underset{a}{\operatorname{argmax}} \{P(1|1,\uparrow)V^\pi(1) + P(2|1,\uparrow)V^\pi(2) + P(3|1,\uparrow)V^\pi(3),$
$P(1|1,\downarrow)V^\pi(1) + P(2|1,\downarrow)V^\pi(2) + P(3|1,\downarrow)V^\pi(3)\}$

$= \underset{a}{\operatorname{argmax}} \{\frac{3}{4}V^\pi(1) + \frac{1}{4}V^\pi(2),\ \frac{1}{4}V^\pi(1) + \frac{3}{4}V^\pi(2)\}$

$= \downarrow$ 「Since second term is larger (has more weight on $V^\pi(2)$」

Similarly



$$\pi'(2) = \arg\max_a \left\{ \tfrac{1}{2}V^\pi(1) + \tfrac{1}{2}V^\pi(2),\ \tfrac{1}{2}V^\pi(2) + \tfrac{1}{2}V^\pi(3) \right\}$$

$= \uparrow$ 「Since $V^\pi(1) > V^\pi(3)$」

$$\pi'(3) = \arg\max_a \left\{ \tfrac{3}{4}V^\pi(2) + \tfrac{1}{4}V^\pi(3),\ \tfrac{1}{4}V^\pi(2) + \tfrac{3}{4}V^\pi(3) \right\}$$

$= \uparrow$ 「First term has "more" $V^\pi(2)$, which is larger」

3.(a) $V^\pi(s) = R(s) + \gamma \sum_{s'} P(s'|s,a) V^\pi(s')$ 「BE」

$$= R(s) + \gamma\left(P(s|s,a)V^\pi(s) + P(s+1|s,a)V^\pi(s+1)\right)$$

$$= R(s) + \gamma\left(\tfrac{2}{3}V^\pi(s) + \tfrac{1}{3}V^\pi(s+1)\right)$$

$$= R(s) + \tfrac{2}{3}\gamma V^\pi(s) + \tfrac{1}{3}\gamma V^\pi(s+1)$$

$$\Rightarrow V^\pi(s) = \frac{s}{1-\frac{2}{3}\gamma} + \frac{\frac{1}{3}\gamma}{1-\frac{2}{3}\gamma} V^\pi(s+1)$$

(b) Write $\beta = \dfrac{s}{1-\frac{2}{3}\gamma} = \dfrac{s}{\frac{3-2\gamma}{3}} = \dfrac{3s}{3-2\gamma}$

$\lambda = \dfrac{\frac{1}{3}\gamma}{1-\frac{2}{3}\gamma} = \dfrac{\gamma}{3-2\gamma}$, so $V^\pi(s) = \beta + \lambda V^\pi(s+1)$

If $V^\pi(s) = as+b \quad \forall s \in \{0, 1, \cdots\}$,

then $V^\pi(s+1) = a(s+1)+b$,

and $V^\pi(s) = \beta + \lambda V^\pi(s+1)$

$$= \beta + \lambda[a(s+1)+b]$$

$$= \beta + \lambda a(s+1) + \lambda b$$

$$\Rightarrow as+b = \frac{3s}{3-2\gamma} + \frac{\gamma a(s+1)}{3-2\gamma} + \frac{b\gamma}{3-2\gamma}$$

$$\Rightarrow (as+b)(3-2\gamma) = 3s + \gamma a(s+1) + b\gamma, \ \forall s$$

Since this is true for all s, then it is true for $s = 1, 2$,

so $\begin{cases} (a+b)(3-2\gamma) = 3 + 2\gamma a + b\gamma \\ (2a+b)(3-2\gamma) = 6 + 3\gamma a + b\gamma \end{cases}$

solve for a, b we have

$$\begin{cases} a = \dfrac{1}{1-\gamma} \\ b = \dfrac{\gamma}{3(\gamma-1)^2} \end{cases}$$

4. See code

5. $\Delta_k = \max_s |V_k(s) - V^\pi(s)|$

$$= \max_s |(R(s) + \gamma \sum_{s'} P(s'|s,\pi(s)) V_{k-1}(s'))$$
$$- (R(s) + \gamma \sum_{s'} P(s'|s,\pi(s)) V^\pi(s'))|$$
$$= \gamma \max_s |\sum_{s'} P(s'|s,\pi(s)) V_{k-1}(s') - \sum_{s'} P(s'|s,\pi(s)) V^\pi(s')|$$
$$= \gamma \max_s |\sum_{s'} [P(s'|s,\pi(s))(V_{k-1}(s') - V^\pi(s'))]|$$
$$= \gamma \max_{s,s'} |\sum_{s'} (V_{k-1}(s') - V^\pi(s'))|$$ 「Choose the largest weight」
$$= \gamma \Delta_{k-1}$$

Since $\gamma < 1$, we have $\Delta_k < \gamma \Delta_{k-1}$.

So $k \to \infty \Rightarrow \Delta_k \to 0$, i.e. $\lim_{k\to\infty} V_k(s) = V^\pi(s)$

6. (a) $\sum_{k=1}^{\infty} \alpha_k = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$,

this is the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n^p}$.

and $\sum \frac{1}{n^p}$ converges if $p > 1$,

diverges if $p \le 1$

「This proof can be found in page 62 of principles of mathematical analysis by Rudin」

So $\sum_{k=1}^{\infty} \alpha_k = \infty$ diverges

$\sum_{k=1}^{\infty} \alpha_k^2 < \infty$ converges

(b) Base: $\mu_1 = \mu_0 + \alpha_1 (X_1 - \mu_0)$

$$= \alpha_1 X_1 = X_1$$

If $\mu_k = \frac{1}{k}(X_1 + \cdots + X_k)$

$$\mu_{k+1} = \mu_k + \alpha_{k+1}(X_{k+1} - \mu_k)$$
$$= \frac{1}{k}\sum_{i=1}^{k} X_i + \frac{X_{k+1} - \frac{1}{k}\sum_{i=1}^{k} X_i}{k+1}$$
$$= \frac{X_{k+1}}{k+1} + \left[\frac{1}{k} - \frac{1}{k(k+1)}\right]\sum_{i=1}^{k} X_i$$
$$= \frac{X_{k+1}}{k+1} + \frac{1}{k+1}\sum_{i=1}^{k} X_i = \frac{1}{k+1}\sum_{i=1}^{k+1} X_i$$