Reinforcement Learning

HW1 Part 1

309505018

郭俊廷

Problem 1

(a) (1) Show $V_*(s) = \max_a Q_*(s,a)$

by definition

$$V^\pi(s) = E_\pi[G_t \mid S_t = s]$$

$$Q^\pi(s,a) = E_\pi[G_t \mid S_t = s, a_t = a]$$

$$\Rightarrow V^\pi(s) = \sum_{a \in A} \pi(a|s) Q^\pi(s,a)$$

$$= E[Q^\pi(s,a)] \le \max Q^\pi(s,a), \ \forall \pi.$$

$\therefore V_*(s) \le \max_a Q_*(s,a)$

If $V_*(s) < \max_a Q_*(s,a)$, then

$\exists \hat{\pi} = \arg\max Q_*(s,a)$

s.t. $V^{\hat{\pi}}(s) > V^*(s)$, which is impossible since π^* is optimal.

$\therefore V_*(s) = \max_a Q_*(s,a)$ #

(2) Show $Q_*(s,a) = R_s^a + \gamma \sum_{s'} P_{ss'}^a V_*(s')$

by definition

$$Q^\pi(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V^\pi(s')$$

$$\Rightarrow Q_*(s,a) = \max_\pi Q^\pi(s,a) = \max\{R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V^\pi(s')\}$$

$$= R_s^a + \max_\pi \{\gamma \sum_{s' \in S} P_{ss'}^a V^\pi(s')\}$$

$$= R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \max_\pi V^\pi(s')$$

$$= R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_*(s')$$

$\therefore Q_*(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_*(s')$ #

(b) show $\|T^*Q - T^*Q'\| \le \gamma \|Q - Q'\|_\infty$

$$\|T^*Q - T^*Q'\|_\infty = \max_{s,a} \left| [T^*Q](s,a) - [T^*Q'](s,a) \right|$$

$$= \max_{s,a} \left| \{R_s^a + \gamma \sum_{s'} P_{ss'}^a \max_{a'} Q(s',a')\} - \{R_s^a + \gamma \sum_{s'} P_{ss'}^a \max_{a'} Q'(s',a')\} \right|$$

$$= \gamma \max_{s,a} \left| \sum_{s'} P_{ss'}^a (\max_{a'} Q(s',a') - \max_{a'} Q'(s',a')) \right|$$

$$\le \gamma \cdot \max_{s,a} \left| \sum_{s'} P_{ss'}^a \max_{a'} (Q(s',a') - Q'(s',a')) \right|$$

$$\le \gamma \max_{s,a} \left| \sum_{s'} P_{ss'}^a \max_{s'',a'} (Q(s'',a') - Q'(s'',a')) \right|$$

$$= \gamma \max_{s,a} \left| \max_{s'',a'} (Q(s'',a') - Q'(s'',a')) \right|$$

$$= \gamma \max_{s'',a'} \left| Q(s'',a') - Q'(s'',a') \right|$$

$$= \gamma \|Q - Q'\|_\infty$$

Therefore, T^* is a γ-contaction operator #

Problem 2

(a) Let U, V be two random variables with CDF F_U, F_V, respectively

Let A be a random variable independant of U and V.

Without loss of generally, U, V, A are assumed to be continuous random variables, with corresponding PDF denoted by f_U, f_V, f_A.

consider any joint PDF $f_{UV}(u,v)$ that has the following marginal PDFs:

$$f_U(u) = \int_R f_{UV}(u,v)\,dv \quad — ①$$

$$f_V(v) = \int_R f_{UV}(u,v)\,du \quad — ②$$

(i)

$$f_{aU}(u) = \int_R f_{aUV}(aU, V)\,dv$$

$$f_{aV}(v) = \int_R f_{UaV}(U, aV)\,du$$

$$f_{aU,aV}(x,y) = \int_R\int_R |a|\, f_{U,V}(u,v)\,du\,dv$$

by definition, $d_p(F_U, F_V) = \inf \|U-V\|_p = \inf \|U-V\|$

$$\int_R\int_R |a| f_{UV}(u,v)\,du\,dv \le |a| \int_R\int_R f_{UV}(u,v)\,du\,dv = |a|\, d_p(U,V), \text{ for any } a \in R$$

(ii)

since A is independent of U, V, by the convolution theorem, we know the PDFs of $(U+A)$ and $(V+A)$

$$f_{U+A}(z) = \int_R f_U(z-a) f_A(a)\,da$$

$$f_{V+A}(z) = \int_R f_V(z-a) f_A(a)\,da.$$

$$f_{U+A,V+A}(x,y) := \int_R f_{UV}(x-a, y-a) f_A(a)\,da$$

$$\int_R\int_R |u-v| f_{UV}(u,v)\,du\,dv = \iint_R |x-y| \cdot \hat{f}_{U+A,V+A}(x,y)\,dx\,dy$$

by definition, $d_p(F_U, F_V) := \inf_{(U,V)} \|U-V\|_p = \inf_{(U,V)} |U-V|$

therefore, for any $\varepsilon > 0$, there must exit a joint PDF $f_{UV}(u,v)$

$$\int_R \int_R |u-v| f_{uv}(u,v)\,du\,dv < d_p(F_U, F_V) + \varepsilon$$

$$\int_R \hat{f}_{U+A,V+A}(x,y)\,dy = \int_R \left(\int_R f_{UV}(x-a, y-a) f_A(a)\,da \right) dy$$

$$= \int_R \left(\int_R f_{UV}(x-a, y-a) f_A(a)\,dy \right) da$$

$$= \int_R f_A(a) f_U(x-a)\,da$$

$$= f_{U+A}(x)$$

$$\int_R f_{U+A,V+A}(x,y)\,dx = \int_R \left(\int_R f_{U,V}(x-a, y-a) f_A(a)\,da \right) dx$$

$$= \int_R \left(\int_R f_{UV}(x-a, y-a) f_A(a)\,dx \right) da$$

$$= \int_R f_A(a) f_V(y-a)\,da$$

$$= f_{V+A}(y)$$

we define $\hat{f}^{(\varepsilon)}_{U+A,V+A}(x,y) := \int_R f^{(\varepsilon)}_{UV}(x-a, y-a) f_A(a)\,da$

$$\int_R \int_R |x-y| \hat{f}_{U+A,V+A}(x,y)\,dx\,dy = \int_R \int_R |u-v| f_{UV}(u,v)\,du\,dv < d_p(F_U, F_V) + \varepsilon$$

for any $\varepsilon > 0$, then we can conclude that

$$d_p(F_{U+A}, F_{V+A}) \le d_p(F_U, F_V)$$

(iii)

$f_{QU}(u) = \int_{\mathbb{R}} f_{QU,V}(QU, V)\, dv$

$f_{QV}(v) = \int_{\mathbb{R}} f_{U,QV}(U, QV)\, du$

$f_{QU,QV}(x,y) = \int_{\mathbb{R}}\int_{\mathbb{R}} |q|\, f_{U,V}(u,v)\, du\, dv$

by definition, $d_p(F_U, F_V) = \inf \|U-V\|_p = \inf \|U-V\|$

$\int_{\mathbb{R}}\int_{\mathbb{R}} |q|\, f_{UV}(u,v)\, du\, dv \leq |q| \int_{\mathbb{R}}\int_{\mathbb{R}} f_{UV}(u,v)\, du\, dv = |q|\, d_p(U,V)$

for any $q \in R$.

(b) Show B^π is a γ-contraction operator in $\bar{d}_p$

consider Z_1, Z_2, by definition

$$\bar{d}_p(B^\pi Z_1, B^\pi Z_2) = \sup_{x,a} d_p(B^\pi Z_1(x,a), B^\pi Z_2(x,a)) \quad \text{--- ①}$$

by the property of d_p

$$d_p(B^\pi Z_1, B^\pi Z_2) = d_p(R(x,a) + \gamma P^\pi Z_1(x,a), R(x,a) + \gamma P^\pi Z_2(x,a))$$

$$\leq \gamma\, d_p(P^\pi Z_1(x,a), P^\pi Z_2(x,a))$$

$$\leq \gamma \sup_{x',a'} d_p(Z_1(x',a'), Z_2(x',a'))$$

(by definition given in original paper:

$P^\pi Z(x,a) := Z(x', A')$, $x' \sim p(\cdot|x,a)$, $A' \sim \pi(\cdot|x')$)

combine with ①

$$\bar{d}_p(B^\pi Z_1, B^\pi Z_2) = \sup_{x,a} d_p(B^\pi Z_1(x,a), B^\pi Z_2(x,a))$$

$$\leq \gamma \sup_{x',a'} d_p(Z_1(x',a'), Z_2(x',a'))$$

$$= \gamma\, \bar{d}_p(Z_1, Z_2)$$

Therefore, $B^\pi : Z \to Z$ is a γ-contraction operator in $\bar{d}_p$. #