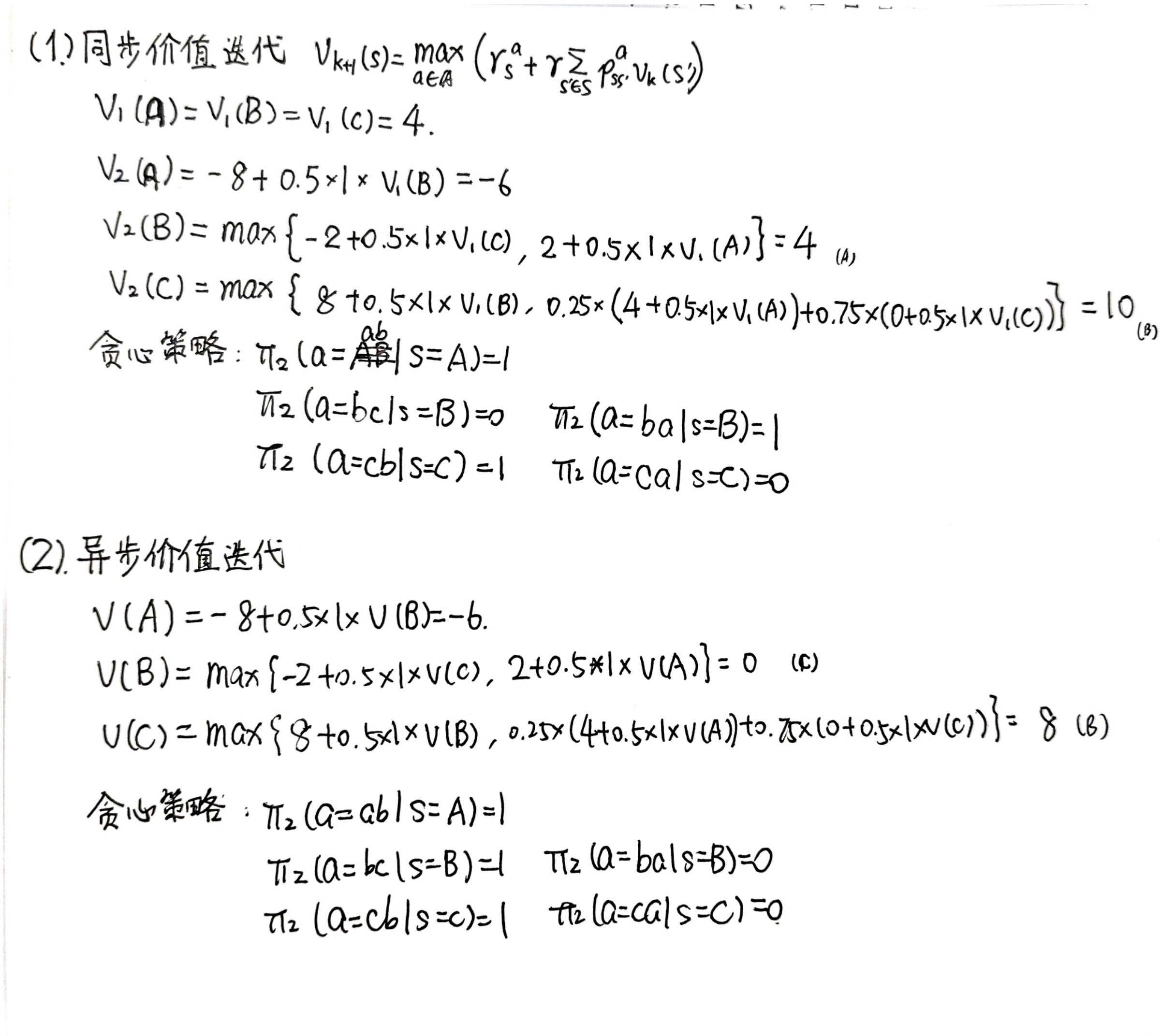
人工智能原理-作业7

Author: **夏弘宇** 2023011004

T1



(1) 同步价值迭代 $V_{k+1}(s)=\max\limits_{a\in A}\left(r_s^a+\gamma\sum\limits_{s'\in S}P_{ss'}^a V_k(s')\right)$

$V_1(A)=V_1(B)=V_1(C)=4.$

$V_2(A)=-8+0.5\times1\times V_1(B)=-6$

$V_2(B)=\max\{-2+0.5\times1\times V_1(C),\ 2+0.5\times1\times V_1(A)\}=4\ _{(A)}$

$V_2(C)=\max\{8+0.5\times1\times V_1(B),\ 0.25\times(4+0.5\times1\times V_1(A))+0.75\times(0+0.5\times1\times V_1(C))\}=10_{(B)}$

贪心策略：$\pi_2(a=\overset{ab}{\cancel{AB}}|s=A)=1$

$\pi_2(a=bc|s=B)=0$ $\pi_2(a=ba|s=B)=1$

$\pi_2(a=cb|s=C)=1$ $\pi_2(a=ca|s=C)=0$

(2). 异步价值迭代

$V(A)=-8+0.5\times1\times V(B)=-6.$

$V(B)=\max\{-2+0.5\times1\times V(C),\ 2+0.5*1\times V(A)\}=0\ \ (C)$

$V(C)=\max\{8+0.5\times1\times V(B),\ 0.25\times(4+0.5\times1\times V(A))+0.75\times(0+0.5\times1\times V(C))\}=8\ \ (B)$

贪心策略：$\pi_2(a=ab|s=A)=1$

$\pi_2(a=bc|s=B)=1$ $\pi_2(a=ba|s=B)=0$

$\pi_2(a=cb|s=C)=1$ $\pi_2(a=ca|s=C)=0$

T2

(1) $V_\pi(S) = \Sigma_{a\in A} \pi(a|s)\left(r_s^a + r \sum_{s'\in S} P_{ss'}^a V_\pi(s')\right)$

$V_\pi(A) = P_{AA}(R_A + V_\pi(A)) + P_{AB}(R_B + V_\pi(B))$

$V_\pi(B) = P_{BA}(R_A + V_\pi(A)) + P_{Bt} \times 0$

解得 $V_\pi(A) = -1$，$V_\pi(B) = +1$.

(2) 首次访问 $V_A = \frac{1}{2}[(3+2-4+4-3)+(\cancel{-2}+3-3)] = 1$

$V_B = \frac{1}{2}[(-4+4-3)+(-2+3-3)] = -2.5$

每次访问 $V_A = \frac{1}{4}[1-1+2+0] = 0.5$

$V_B = \frac{1}{4}[-3-3-3-2] = -2.75$

T3

(1) 时序差分 $V_{t+1}(S_t)=V_t(S_t)+\alpha(r_{t+1}+\gamma V_t(S_{t+1})-V_t(S_t))$

4-1

0	0	0
0	-0.5	0
0	0	0

1-4

0	-0.75	0
0	-0.5	0
0	0	0

4

0	-0.75	0
0	-0.75	0
0	0	0

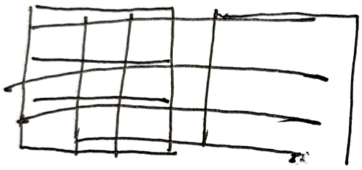
7

0	-0.75	0
0	-0.75	0
0	-0.5	0

terminate

(2) SARSA算法 $q_{t+1}(S_t,a_t)=q_t(S_t,a_t)+\alpha_t(r_{t+1}+\gamma q_t(S_{t+1}a_{t+1})-q_t(S_t a_t))$

4 → 7 → 6 → 3 → terminate



Q(4, 下) ← R + Q(7, 右)

Q(7, 右) ← R + Q(6, 上)

Q(6, 上) ← R + Q(3, 上)

Q(3, 上) ← R.

	1	2	3	4	5	6	7
上	-4	-3	-1	-3	-4	-2	-4
左	-3	-3	-2	-4	-2	-3	-3
下	-4	-3	-4	-3	-2	-3	-4
右	-3	-2	-3	-3	-4	-3	-3