

Задача 1: Автономно возило

а)

#	Податоци	Вредност на тежините(ажурирање) по добивање на податоците
0		1, 0, 0, 0
1	Почетна состојба на сензорите: D=0, S=2 Акција: A Награда: -2 Крајна состојба на сензорите: D=1, S=0	1, -1, 0, 0
2	Почетна состојба на сензорите: D=1, S=0 Акција: B Награда: 0 Крајна состојба на сензорите: D=1, S=0	1, -1, 0.5, 0

Ќе користам линеарна Q-функција.

$$Q(s, a) = w_1 f_1(s, a) + w_2 f_2(s, a) + \dots + w_n f_n(s, a)$$

$$difference = [r + \gamma \max Q(s', a')] - Q(s, a)$$

Иницијални тежини се $w_{AD} = 1$, $W_{AS} = W_{BD} = W_{BS} = 0$

$$\gamma = 1, \alpha = 0.5$$

Епизода 1:

$$s = (D = 0, S = 2)$$

Action: A

$$R = -2$$

$$s' = (D = 1, S = 0)$$

Пресметка на карактеристиките:

$$f_{BD} = f_{BS} = 0 \text{ бидејќи акцијата е A}$$

$$f_{AD} = 0, f_{AS} = 2$$

Пресметка на Q вредности:

$$Q(s, A) = w_{AD} * f_{AD} + w_{AS} * f_{AS} = 1 * 0 + 0 * 2 = 0$$

Q-вредности за новата состојба:

$$f_{AD} = 1, f_{AS} = 0$$

$$Q(s', A) = w_{AD} * f_{AD} + w_{AS} * f_{AS} = 1 * 1 + 0 * 0 = 1$$

$$f_{BD} = 1, f_{BS} = 0$$

$$Q(s', B) = w_{BD} * f_{BD} + w_{BS} * f_{BS} = 0 * 1 + 0 * 0 = 0$$

$$\begin{aligned} Q_{new}(s, A) &= R(s, A, s') + \gamma * \max_{a'} Q(s', a') = \\ &= -2 + \gamma * \max\{Q(s', A), Q(s', B)\} = -2 + 1 * 1 = -1 \\ difference &= Q_{new}(s, A) - Q(s, A) = -1 - 0 = -1 \end{aligned}$$

Ажурирање тежини:

$$w_i = w_i + \alpha * (difference) * (f_i)$$

$$w_{AD} = w_{AD} + \alpha * (difference) * (f_{AD})$$

$$w_{AD} = 1 + 0.5 * (-1) * 0 = 1$$

$$w_{AS} = w_{AS} + \alpha * (difference) * (f_{AS})$$

$$w_{AS} = 0 + 0.5 * (-1) * 2 = -1$$

Епизода 2:

$$s = (D = 1, S = 0)$$

Action: B

$$R = 0$$

$$s' = (D = 1, S = 0)$$

Пресметка на карактеристиките:

$$f_{AD} = f_{AS} = 0 \text{ бидејќи акцијата е } B$$

$$f_{BD} = 1, f_{BS} = 0$$

Пресметка на Q вредности:

$$Q(s, B) = w_{BD} * f_{BD} + w_{BS} * f_{BS} = 0 * 1 + 0 * 0 = 0$$

Q-вредности за новата состојба:

$$f_{AD} = 1, f_{AS} = 0$$

$$Q(s', A) = w_{AD} * f_{AD} + w_{AS} * f_{AS} = 1 * 1 - 1 * 0 = 1$$

$$f_{BD} = 1, f_{BS} = 0$$

$$Q(s', B) = w_{BD} * f_{BD} + w_{BS} * f_{BS} = 0 * 1 + 0 * 0 = 0$$

$$\begin{aligned} Q_{new}(s, B) &= R(s, B, s') + \gamma * \max_{a'} Q(s', a') = \\ &= 0 + \gamma * \max\{Q(s', A), Q(s', B)\} = 0 + 1 * 1 = 1 \\ difference &= Q_{new}(s, B) - Q(s, B) = 1 - 0 = 1 \end{aligned}$$

Ажурирање тежини:

$$w_{BD} = w_{BD} + \alpha * (difference) * (f_{BD})$$

$$w_{BD} = 0 + 0.5 * 1 * 1 = 0.5$$

$$w_{BS} = w_{BS} + \alpha * (difference) * (f_{BS})$$

$$w_{BS} = 0 + 0.5 * 1 * 0 = 0$$

б) Акцијата која ќе ја преземе агентот во оваа состојба е В(закочи). Образложение:
 $s=(D=1, S=1)$

$$w_{AD} = 1, w_{AS} = -1, w_{BD} = 0.5, w_{BS} = 0$$

$$Q(s, A) = w_{AD} * f_{AD} + w_{AS} * f_{AS} = 1 * 1 - 1 * 1 = 0$$

$$Q(s, B) = w_{BD} * f_{BD} + w_{BS} * f_{BS} = 0.5 * 1 + 0 * 1 = 0.5$$

Акцијата е В бидејќи се добива поголема Q-вредност.

Задача 2: Ајде да играме „Мунти 21“

а)

$$V_k(13) = 2$$

$$V_k(s) = 10, \text{ за } s = \{14, 15, \dots, 21\}$$

$$V_k(\text{Изгоре}) = -10$$

$$V_k(\text{Доста}) = 0$$

$$V_{k+1}(12) = ?$$

Од белмановата равенка:

$$V_{k+1}(s) = \max_{s'} [\sum T(s, a, s') * (R(s, a, s') + \gamma V_k(s'))]$$
 (бидејќи гама е 1, не го пишувам во понатамошните пресметки)

$$V_{k+1}(12) = \max_{s'} [\sum T(s, a, s') * (R(s, a, s') + \gamma V_k(s'))]; s = \{14, \dots, 21, J, K, Q, A\}, V_k(\text{изгоре}), V_k(\text{доста})]$$

$$V_{k+1}(12) = \max[\frac{1}{13} * ((8 * 10) + 5 * (-10)), -10] = \max[\frac{1}{13} (80 - 50), -10, 0]$$

$$V_{k+1}(12) = \max[\frac{30}{13}, -10, 0] = 2.31$$

Образложение:

Бидејќи има 13 карти кои може да се извлечат, а претпоставуваме дека тие имаат рамномерна распределба, секоја карта има веројатност од $\frac{1}{13}$ за да биде извлечена. Од тука доаѓа таа бројка во пресметките.

б) Иницијални вредности

s	a	Q(s,a)
19	Влечи	-2
19	Доста	5
20	Влечи	-4
20	Доста	7
21	Влечи	-6
21	Доста	8
Изгоре	Доста	-8

Епизода

s	a	r	s	a	r	s	a	r
19	Влечи	0	21	Влечи	0	Изгоре	Доста	-10

Формула која е потребна за решението:

$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a')]$, ќе ја користам без пишување на гама бидејќи факторот на намалување $\gamma = 1$.

$$\gamma = 1$$

$$\alpha = 0.5$$

Чекор 1:

$$s = 19, a = \text{Влечи}, r = 0, s' = 21$$

$\max_{a'} Q(21, a') = \max(-6, 8) = 8$ Вредностите -6 и 8 се прочитани од табелата со иницијални вредности, кај состојба 21. (соодветно и за последователните чекори).

$$Q(19, \text{Влечи}) = -2$$

Ажурирање на вредноста:

$$Q(19, \text{Влечи}) \leftarrow 0.5 * (-2) + 0.5 * (0 + 8)$$

$$Q(19, \text{Влечи}) \leftarrow -1 + 4$$

$$Q(19, \text{Влечи}) \leftarrow 3$$

Чекор 2:

$$s = 21, a = \text{Влечи}, r = 0, s' = \text{Изгоре}$$

$$\max_{a'} Q(\text{Изгоре}, a') = \max(-8) = -8$$

$$Q(21, \text{Влечи}) = -6$$

Ажурирање на вредноста:

$$Q(21, \text{Влечи}) \leftarrow 0.5 * (-6) + 0.5 * (0 - 8)$$

$$Q(21, \text{Влечи}) \leftarrow -3 - 4$$

$$Q(21, \text{Влечи}) \leftarrow -7$$

Чекор 2:

$s = \text{Изгоре}$, $a = \text{Доста}$, $r = -10$, $s' = \text{Крај}$

Бидејќи е крај на играта $\max_a Q(\text{Изгоре}) = 0$

$$Q(\text{Изгоре}, \text{Доста}) = -8$$

Ажурирање на вредноста:

$$Q(\text{Изгоре}, \text{Доста}) \leftarrow 0.5 * (-8) + 0.5 * (-10 + 0)$$

$$Q(\text{Изгоре}, \text{Доста}) \leftarrow -4 - 5$$

$$Q(\text{Изгоре}, \text{Доста}) \leftarrow -9$$

Бидејќи во епизодата немаме повеќе акции, останатите Q-вредности нема да се променат, дополнително ги означувам како непроменети и во табелата.. Конечната табела со вредности (Односно табела 2 од описот на домашната) е:

s	a	Q(s,a)
19	Влечи	3
19	Доста	5 (непроменето)
20	Влечи	-4 (непроменето)
20	Доста	7 (непроменето)
21	Влечи	-7
21	Доста	8 (непроменето)
Изгоре	Доста	-9

в)

Политика 1:

s	$\pi(s)$
14	Влечи
15	Влечи
16	Влечи
17	Влечи
18	Влечи
19	Доста

$$Q(14, \text{Влечи}) = w_1(-1) + w_2(-1) = -w_1 - w_2$$

$$Q(15, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(16, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(17, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(18, \text{Влечи}) = w_1(1) + w_2(1) = w_1 + w_2$$

$$Q(19, \text{Доста}) = 0$$

Бидејќи секогаш за акција Доста, Q-вредноста е 0, треба тежините да се такви што агентот ќе ја преферира Влечи акцијата.

Па така, $Q(14, \text{Влечи}) = -w_1 - w_2$ треба да е поголемо од $Q(19, \text{Доста}) = 0$, односно:

$$-w_1 - w_2 > 0$$

Од друга страна, за $Q(18, \text{Влечи}) = w_1 + w_2$ да биде подобро од $Q(19, \text{Доста}) = 0$, треба:

$w_1 + w_2 > 0$. Тука се јавува контрадикција, бидејќи $w_1 + w_2 = -(-w_1 - w_2)$, односно е невозможно двата изрази да се поголеми од 0.

Политика 2:

s	$\pi(s)$
14	Влечи
15	Влечи
16	Влечи
17	Влечи
18	Доста
19	Доста

$$Q(14, \text{Влечи}) = w_1(-1) + w_2(-1) = -w_1 - w_2$$

$$Q(15, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(16, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(17, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(18, \text{Доста}) = w_1(0) + w_2(0) = 0$$

$$Q(19, \text{Доста}) = 0$$

За $Q(14, \text{Влечи})$ го имаме истиот израз како и претходно, односно $-w_1 - w_2 > 0$

Тука треба да се провери дали има некакви контрадикции со

$$Q(15, \text{Влечи}) > 0, \text{ односно } w_1 - w_2 > 0$$

Доколку ги средиме изразите:

$$-w_1 - w_2 > 0 \rightarrow w_1 + w_2 \leq 0$$

$$w_1 > w_2$$

Тука немаме контрадикција, бидејќи може да се најде состојба на тежините за кои двете неравенства се задоволени: $w_1 = -2$, $w_2 = -3$, односно неравенствата се задоволени кога $(w_1 < 0 \wedge w_2 < 0) \wedge (|w_1| < |w_2|)$.

Со горенаведеното се покриваат и врските помеѓу $Q(14, \text{Влечи})$ и $Q(16, \text{Влечи})$ и $Q(14, \text{Влечи})$ и $Q(17, \text{Влечи})$.

Од друга страна, $Q(14, \text{Влечи})$ и $Q(18, \text{Влечи})$ Немаат никаква контрадикција бидејќи $Q(18, \text{Влечи})$ е 0 независно од тежините, истото важи и за $Q(19, \text{Доста})$

Политика 3:

s	$\pi(s)$
14	Доста
15	Влечи
16	Влечи
17	Влечи
18	Доста
19	Доста

$$Q(14, \text{Влечи}) = w_1(0) + w_2(0) = 0$$

$$Q(15, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(16, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(17, \text{Влечи}) = w_1(1) + w_2(-1) = w_1 - w_2$$

$$Q(18, \text{Доста}) = w_1(0) + w_2(0) = 0$$

$$Q(19, \text{Доста}) = 0$$

Согласно напишаното за претходната политика, нема никакви контрадицкии бидејќи Q-вредностите кои се 0, се 0 независно од тежините.