Machine, Data & Learning

Monday

24/01/2023

Aayem Ajay Sharma
Roll no: 20221210001

Bellman Equation:

$$U(s) = R(s,a) + \gamma \cdot \max\left[\sum_{s'} P(s'|s,a)\, U(s')\right]$$

Iteration - 0:

U_0	1	2	3
0	0	1	-1
1	0	0	0
2	0	WALL	0
3	0	0	0

Iteration - 1:

$U[0,0] = -0.04 + 0.95 \max[0.7\times0 + 0.15\times0 + 0.15\times0$, up
$0.7\times0 + 0.15\times0 + 0.15\times1$, down
$0.7\times0 + 0.15\times0 + 0.15\times0$, left
$0.7\times1 + 0.15\times0 + 0.15\times0$, right$]$

$= -0.04 + 0.95 \times 0.07$

$= 0.625$

$U[0,1] = 1$

$U[0,2] = -1$

__/__/____

$U[1,0] = -0.04 + 0.95 \max[0.7\times0 + 0.15\times0 + 0.15\times0$, up
$0.7\times0 + 0.15\times0 + 0.15\times0$, down
$0.7\times0 + 0.15\times0 + 0.15\times0$, Left
$0.7\times0 + 0.15\times0 + 0.15\times0$, right$]$

$= -0.04 + 0.95 \times 0$

$= -0.04$

$U[1,1] = -0.04 + 0.95 \max[0.7\times1 + 0.15\times0 + 0.15\times0$, up
$0.7\times0 + 0.15\times0 + 0.15\times0$, down
$0.7\times0 + 0.15\times1 + 0.15\times0$, left
$0.7\times0 + 0.15\times0 + 0.15\times1$, right$]$

$= -0.04 + 0.95 \times 0.7$

$= 0.625$

$U[1,2] = -0.04 + 0.95 \max[0.7\times-1 + 0.15\times0 + 0.15\times0$, up
$0.7\times0 + 0.15\times0 + 0.15\times0$, down
$0.7\times0 + 0.15\times-1 + 0.15\times0$, left
$0.7\times0 + 0.15\times0 + 0.15\times-1$, right$]$

$= -0.04 + 0.95 \times 0$

$= -0.04$

$U[2,0] = -0.04 + 0.95 \times \max[0.7\times0 + 0.15\times0 + 0.15\times0$, up
$0.7\times0 + 0.15\times0 + 0.15\times0$, down
$0.7\times0 + 0.15\times0 + 0.15\times0$, left
$0.7\times0 + 0.15\times0 + 0.15\times0$, right$]$

__/__/____

$U[2,0] = -0.04 + 0.95 \times 0.$

$= -0.04.$

$U[2,1] = \text{WALL}$

$U[2,2] = -0.04 + 0.95 \max[0.7\times0+0.15\times0+0.15\times0, \text{up}$

$0.7\times0+0.15\times0+0.15\times0, \text{down}$

$0.7\times0+0.15\times0+0.15\times0, \text{left}$

$0.7\times0+0.15\times0+0.15\times0, \text{right}]$

$= -0.04 + 0.95 \times 0$

$= -0.04$

$U[3,0] = -0.04 + 0.95 \max[0.7\times0+0.15\times0+0.15\times0, \text{up}$

$0.7\times0+0.15\times0+0.15\times0, \text{left}$

$0.7\times0+0.15\times0+0.15\times0, \text{down}$

$0.7\times0+0.15\times0+0.15\times0, \text{right}]$

$= -0.04 + 0.95 \times 0$

$= -0.04$

$U[3,1] = -0.04 + 0.95 \max[0.7\times0+0.15\times0+0.15\times0, \text{up}$

$0.7\times0+0.15\times0+0.15\times0, \text{down}$

$0.7\times0+0.15\times0+0.15\times0, \text{left}$

$0.7\times0+0.15\times0+0.15\times0, \text{right}]$

$U[3,1] = -0.04 + 0.95 \times 0$

$= -0.04.$

__/__/____

$$U[3,2] = -0.04 + 0.95 \max[0.7\times0 + 0.15\times0 + 0.15\times0, \text{up}$$
$$0.7\times0 + 0.15\times0 + 0.15\times0, \text{down}$$
$$0.7\times0 + 0.15\times0 + 0.15\times0, \text{left}$$
$$0.7\times0 + 0.15\times0 + 0.15\times0, \text{right}]$$
$$= -0.04 + 0.95\times0$$
$$= -0.04$$

$$\therefore U_1 = \begin{bmatrix} 0.625 & 1 & -1 \\ -0.04 & 0.625 & -0.04 \\ -0.04 & \text{WALL} & -0.04 \\ -0.04 & -0.04 & -0.04 \end{bmatrix}$$

Iteration 2 :

$$U[0,0] = -0.04 + 0.95 \max[(\text{up}) 0.7\times0.625 + 0.15\times0.625 + 0.15\times$$
$$(\text{down}) \ 0.7\times0.04 + 0.15\times0.625 + 0.15\times1$$
$$(\text{left}) \ 0.7\times0.625 + 0.15\times0.625 + 0.15\times-0.04$$
$$(\text{Right}) \ 0.7\times1 * 0.15\times0.625 + 0.15\times-0.04]$$

$$U[0,0] = -0.04 + 0.95(0.7\times1 + 0.15\times0.625 + 0.15\times-0.04)$$
$$= -0.04 + 0.95\times0.788$$
$$= -0.04 + 0.748$$
$$= 0.708$$

__/__/____

$U[0,1] = +1$ (Terminating states)

$U[0,2] = -1$ (Terminating state)

$U[1,1] = -0.04 + 0.95 \times \max$ of

1) $0.7 \times 1 + 0.15 \times -0.04 + 0.15 \times -0.04$, up
2) $0.7 \times 0.625 + 0.15 \times -0.04 + 0.15 \times -0.04$, down
3) $0.7 \times -0.04 + 0.15 \times 1 + 0.15 \times 0.625$ left
4) $0.7 \times -0.04 + 0.15 \times 1 + 0.15 \times 0.625$, right

$= -0.04 + 0.95(0.7 \times 1 + 0.15 \times -0.04 + 0.15 \times -0.04)$
$= -0.04 + 0.95 \times 0.688$
$= -0.04 + 0.654$
$= 0.614.$

$\boxed{U[1,1] = 0.614}$

$U[1,2] = -0.04 + 0.95 \times \max$ of

1) $0.7 \times -1 + 0.15 \times 0.625 + 0.15 \times -0.04$, up
2) $0.7 \times -0.04 + 0.15 \times 0.625 + 0.15 \times -0.04$, down
3) $0.7 \times 0.625 + 0.15 \times -1 + 0.15 \times -0.04$, left
4) $0.7 \times -0.04 + 0.15 \times -1 + 0.15 \times -0.04$, right

$U[1,2] = -0.04 + 0.95 \times (0.7 \times 0.625 + 0.15 \times -1 + 0.15 \times -0.04)$
$= -0.04 + 0.95 \times 0.282$

__/__/____

$= -0.04 + 0.267$

$= 0.227$

$U[1,2] = 0.227$

$U[1,0] = -0.04 + 0.95 \times \max$ of

(up) 1) $0.7 \times 0.625 + 0.15 \times -0.04 + 0.15 \times 0.625$

(down) 2) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times 0.625$

(left) 3) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times 0.625$

(right) 4) $0.7 \times 0.625 + 0.15 \times -0.04 + 0.15 \times 0.625$

$U[1,0] = -0.04 + 0.95 \times (0.7 \times 0.625 + 0.15 \times -0.04 + 0.15 \times 0.625)$

$= -0.04 + 0.95 \times 0.525$

$= -0.04 + 0.499$

$= 0.459.$

$U[1,0] = 0.459$

$U[2,0] = -0.04 + 0.95 *$ max of

(up) 1) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(down) 2) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(left) 3) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(right) 4) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04.$

__/__/____

$U[2,0] = -0.04 + 0.95 \times (0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04)$

$= -0.04 + 0.95 \times -0.04$

$= 1.95 \times -0.04$

$= -0.078$

$U[2,0] = -0.078$

$U[2,1] = WALL$

$U[2,2] = -0.04 + 0.95 \times$ max of

(up) 1) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(down) 2) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(left) 3) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(right) 4) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

$U[2,2] = -0.04 + 0.95 \times (0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04)$

$= -0.04 + 0.95 \times -0.04$

$= 1.95 \times -0.04$

$= -0.078$

$U[2,2] = -0.078$

$U[3,0] = -0.04 + 0.95 \times$ max of

(up) 1) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(down) 2) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(left) 3) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

(right) 4) $0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04$

__/__/____

$V[3,0] = -0.04 + 0.95[0.7x-0.04+0.15x-0.04+0.15x-0.04)$

$= -0.04 + 0.95x-0.04$

$= -0.04 * 1.95$

$= -0.078$

$V[3,1] = -0.04 +$ max of

(up) 1) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

(down) 2) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

(left) 3) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

(right) 4) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

$V[3,1] = -0.04 + 0.95 x(0.7x-0.04+0.15x-0.04+0.15x-0.04))$

$= -0.04 + 0.95x-0.04$

$= 1.95x-0.04$

$= -0.078$

$V[3,1] = -0.078.$

$V[3,2] = -0.04 + 0.95 x$ max of

(up) 1) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

(down) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

(left) 3) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

(right) 4) $0.7x-0.04+0.15x-0.04+0.15x-0.04$

$\therefore V[3,2] = -0.04 + 0.95x(0.7x-0.04+0.15x-0.04+0.15x-0.04)$

__/__/____

$$\therefore U[3,2] = -0.04 + 0.95 \times -0.04$$
$$= 1.95 \times -0.04$$
$$= -0.078$$
$$U[3,2] = -0.078.$$

$$\therefore U_2 = \begin{bmatrix} 0.708 & 1 & -1 \\ 0.459 & 0.614 & 0.227 \\ -0.078 & \text{WALL} & -0.078 \\ -0.078 & -0.078 & -0.078 \end{bmatrix}$$

The values of the grid match with output of the code.

* [U[1,1] = 0.614 in the dry run, but in the code it's U[1,1] = 0.613. This is probably due to how Python handles & approximates floats].