MDL Assignment3

Roll : 2021101006

2/ Given grid world

	Reward: 1	Penalty: -1
	wall	
Start		

Grid =
[[0, 1, -1],
[0, 0, 0],
[0, 'wall', 0],
[0, 0, 0]]

Iteration-0: Initial utility values of all states are zero except reward and penalty.

Assumption: utility value of "wall" is zero

	0	1	2
0	0	1	-1
1	0	0	0
2	0	0	0
3	0	0	0

⇒ values in cells represents utility of corresponding cell

* $$U_{t+1}(I) = \max_{A} \left(R(I,A) + \gamma \sum_{J} P(J/I,A)\, U_t(J) \right)$$ — (a)

where, I is current cell, J is next state,
A is action performed

Given, discount factor = $\gamma = 0.95$

(a) Step cost = $R(I,A) = -0.04 \; \forall I, A$

* Given, probability of going in
the direction of an action = 0.7
direction perpendicular to action = 0.15

Here 0.7 + 0.15 + 0.15 = 1 ✓

Iteration-1

From ②,

$$V_1(0,0) = \max \begin{cases} -0.04 + 0.95[0.7\times0 + 0.15\times1 + 0.15\times0] \\ -0.04 + 0.95[0.7\times1 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times1] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} 0.1025 \\ 0.625 \\ 0.1025 \\ -0.04 \end{cases} = 0.625 \Rightarrow V_1(0,0) = 0.625$$

$V_1(0,1) = 1$ [From question]

$V_1(0,2) = -1$ [From question]

From ②,

$$V_1(1,0) = \max \begin{cases} -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} -0.04 \\ -0.04 \\ -0.04 \\ -0.04 \end{cases} = -0.04 \Rightarrow V_1(1,0) = -0.04$$

$$V_1(1,1) = \max \begin{cases} -0.04 + 0.95[0.7\times1 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times1] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times1 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} 0.625 \\ 0.1025 \\ -0.04 \\ 0.1025 \end{cases} = 0.625 \Rightarrow V_1(1,1) = 0.625$$

$$V_1(1,2) = \max \begin{cases} -0.04 + 0.95[0.7\times(-1) + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times(-1)] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times(-1) + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} -0.705 \\ -0.1825 \\ -0.04 \\ -0.1825 \end{cases} = -0.04 \Rightarrow V_1(1,2) = -0.04$$

$$V_1(2,0) = \max \begin{cases} -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} -0.04 \\ -0.04 \\ -0.04 \\ -0.04 \end{cases} = -0.04 \Rightarrow V_1(2,0) = -0.04$$

From question and assumption,

$$V_1(2,1) = 0$$

$$U_1(2,2) = \max \begin{cases} -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} -0.04 \\ -0.04 \\ -0.04 \\ -0.04 \end{cases} = -0.04 \Rightarrow U_1(2,2) = -0.04$$

$$U_1(3,0) = \max \begin{cases} -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} -0.04 \\ -0.04 \\ -0.04 \\ -0.04 \end{cases} = -0.04 \Rightarrow U_1(3,0) = -0.04$$

$$U_1(3,1) = \max \begin{cases} -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} -0.04 \\ -0.04 \\ -0.04 \\ -0.04 \end{cases} = -0.04 \Rightarrow U_1(3,1) = -0.04$$

$$U_1(3,2) = \max \begin{cases} -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \\ -0.04 + 0.95[0.7\times0 + 0.15\times0 + 0.15\times0] \end{cases}$$

$$= \max \begin{cases} -0.04 \\ -0.04 \\ -0.04 \\ -0.04 \end{cases} = -0.04 \Rightarrow U_1(3,2) = -0.04$$

∴ Grid with corresponding utility values of each cell after first iteration is:

0.625	1	-1
-0.04	0.625	-0.04
-0.04	0	-0.04
-0.04	-0.04	-0.04

— (I)

Iteration-2:

From (a),

$U_2(0,0)$

$$= \max \begin{cases} -0.04 + 0.95[0.7 \times 0.625 + 0.15 \times 1 + 0.15 \times 0.625] \\ -0.04 + 0.95[0.7 \times 1 + 0.15 \times (-0.04) + 0.15 \times 0.625] \\ -0.04 + 0.95[0.7 \times (-0.04) + 0.15 \times 0.625 + 0.15 \times 1] \\ -0.04 + 0.95[0.7 \times 0.625 + 0.15 \times 0.625 + 0.15 \times (-0.04)] \end{cases}$$

$$= \max \begin{cases} 0.6071875 \\ 0.7083625 \\ 0.2181625 \\ 0.4589875 \end{cases} = 0.7083625$$

$$\Rightarrow U_2(0,0) = 0.7083625$$

From question,

$U_2(0,1) = 1$

$U_2(0,2) = -1$

From (a),

$$V_2(1,0) = \max \begin{cases} -0.04 + 0.95[0.7 \times 0.625 + 0.15 \times 0.625 + 0.15 \times (-0.04)] \\ -0.04 + 0.95[0.7 \times 0.625 + 0.15 \times (-0.04) + 0.15 \times 0.625] \\ -0.04 + 0.95[0.7 \times (-0.04) + 0.15 \times (-0.04) + 0.15 \times 0.625] \\ -0.04 + 0.95[0.7 \times (-0.04) + 0.15 \times 0.625 + 0.15 \times (-0.04)] \end{cases}$$

$$= \begin{cases} 0.4589875 \\ 0.4589875 \\ 0.0167625 \\ 0.0167625 \end{cases} = 0.4589875 \Rightarrow V_2(1,0) = 0.4589875$$

$V_2(1,1)$.

$$= \max \begin{cases} -0.04 + 0.95[0.7 \times 1 + 0.15 \times (-0.04) + 0.15 \times (-0.04)] \\ -0.04 + 0.95[0.7 \times (-0.04) + 0.15 \times 0.625 + 0.15 \times 1] \\ -0.04 + 0.95[0.7 \times 0.625 + 0.15 \times (-0.04) + 0.15 \times (-0.04)] \\ -0.04 + 0.95[0.7 \times (-0.04) + 0.15 \times 1 + 0.15 \times 0.625] \end{cases}$$

$$= \begin{cases} 0.6136 \\ 0.1649625 \\ 0.364225 \\ 0.1649625 \end{cases} = 0.6136 \Rightarrow V_2(1,1) = 0.6136$$

$V_2(1,2)$

$$= \max \begin{cases} -0.04 + 0.95[0.7 \times (-1) + 0.15 \times (-0.04) + 0.15 \times 0.625] \\ -0.04 + 0.95[0.7 \times (-0.04) + 0.15 \times (0.04) + 0.15 \times (-1)] \\ -0.04 + 0.95[0.7 \times (-0.04) + 0.15 \times 0.625 + 0.15 \times (-0.04)] \\ -0.04 + 0.95[0.7 \times 0.625 + 0.15 \times (-1) + 0.15 \times (-0.04)] \end{cases}$$

$$= \max \begin{cases} -0.6216375 \\ -0.2148 \\ 0.0167625 \\ 0.227425 \end{cases} = 0.227425$$

$$\Rightarrow V_2(1,2) = 0.227425$$

$V_2(2,0)$

$$= \max \begin{cases} -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)], \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \end{cases}$$

$$= \max \begin{cases} -0.078 \\ -0.078 \\ -0.078 \\ -0.078 \end{cases} = -0.078 \Rightarrow V_2(2,0) = -0.078$$

From question and assumption, $V_2(2,1) = 0$

From (a),

$V_2(2,2)$

$$= \max \begin{cases} -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \end{cases}$$

$$= \max \begin{cases} -0.078 \\ -0.078 \\ -0.078 \\ -0.078 \end{cases} = -0.078 \Rightarrow V_2(2,2) = -0.078$$

$V_2(3,0)$

$$= \max \begin{cases} -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \end{cases}$$

$$= \max \begin{cases} -0.078 \\ -0.078 \\ -0.078 \\ -0.078 \end{cases} = -0.078 \Rightarrow V_2(3,0) = -0.078$$

$U_2(3,1)$

$$= \max \begin{cases} -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \end{cases}$$

$$= \max \begin{cases} -0.078 \\ -0.078 \\ -0.078 \\ -0.078 \end{cases} = -0.078 \Rightarrow U_2(3,1) = -0.078$$

$U_2(3,2)$

$$= \max \begin{cases} -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \\ -0.04 + 0.95[0.7(-0.04) + 0.15(-0.04) + 0.15(-0.04)] \end{cases}$$

$$= \max \begin{cases} -0.078 \\ -0.078 \\ -0.078 \\ -0.078 \end{cases} = -0.078 \Rightarrow U_2(3,2) = -0.078$$

∴ Grid with corresponding utility values of each cell after second iteration is:

0.7083625	1	-1
0.4589875	0.6136	0.227425
-0.078	0	-0.078
-0.078	-0.078	-0.078

— (II)

From (I), (II),
Grids obtained after first and second iterations matches with the corresponding output from my code.