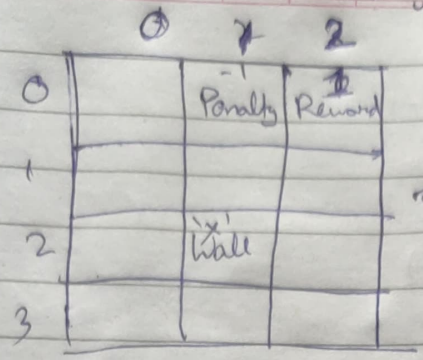
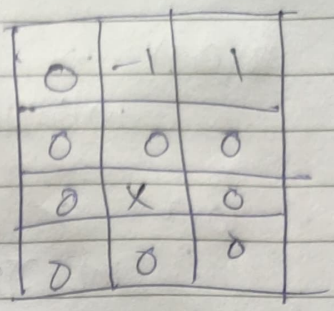
MDL Assignment 3.2

Initialize other values to 0





$\gamma = 0.7$ for 2 iterations.

We take utility of itself, if any cell is out of matrix or it's wall

$(r,c) = (0,0)$

$$\text{Policy}(I) = \text{argmax}_A \left(C(I,A) + \gamma \sum_J P(J/I,A)\, U_t(J) \right)$$

A = Down

$$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(-1) \right]$$

(U itself)

$$= -0.1825$$

A = UP

$$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(-1) \right]$$

(U itself)

$$= -0.1825$$

A = Left

$$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$$

(U itself; $+\left(\frac{1-0.7}{2}\right)(0)$ → $(1,0)$)

$$= -0.04$$

A = Right

$$= -0.04 + (0.95)\left[(0.7)(-1) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$$

$$= -0.705$$

$M[0][0] = \max(-0.1825, -0.1825, -0.04, -0.705)$

$M[0][0] = -0.04$

i.e left.

(0,1) → Penalty (-1)

$\therefore M[0][1] = -1$

(0,2) → Reward (1)

$\therefore M[0][2] = 1$

~~(2,0)~~

let's do for (1,0)

Policy $[I] = \text{argmax}_A \left[C(I,A) + \gamma \sum_J P(J|I,A) \cdot U_t(I) \right)$

A = down

$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$

$= -0.04$

A = UP

$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$

$= -0.04$

A = Left

$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$

$= -0.04$

A = Right

$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$

$= -0.04$

$M[1][0] = \max(-0.04, -0.04, -0.04, -0.04)$

$= -0.04$

Page No.
Date

Let's do for (1,2)

$$\text{Policy}(I) = \underset{A}{\text{argmax}} \left[([I,A] + \gamma \sum_{J} P(J|I,A)\, U_t(J) \right]$$

A = down:

$$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$$

$$= -0.04$$

A = UP

$$= -0.04 + (0.95)\left[(0.7)(1) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0) \right]$$

$$= 0.625$$

A = Left

$$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(1) + \left(\frac{1-0.7}{2}\right)(0) \right]$$

$$= 0.1025$$

A = Right

$$= -0.04 + (0.95)\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(1) + \left(\frac{1-0.7}{2}\right)(0) \right]$$

$$= 0.1025$$

$$M[1][2] = \max(-0.04, 0.625, 0.1025, 0.1025)$$

$$= 0.625$$

(0,1)

max

$= -0.04 + 0.95$

$= \begin{cases} \text{Down} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{up} & -0.04 + 0.95\left[(0.7)(-1) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{left} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(-1) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{max Right} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(+1)\right] \end{cases}$

$= -0.04$

(2,0)

$\begin{cases} \text{Down} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{up} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{left} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{max Right} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \end{cases}$

$= -0.04$

(2,1) Wall

(2,2)

$\begin{cases} \text{Down} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{up} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{left} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \\ \text{max Right} & -0.04 + 0.95\left[(0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right] \end{cases}$

$= -0.04$

(3,0)

$$\max \begin{cases} \text{Down} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{up} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{left} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{Right} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \end{cases}$$

$= -0.04$

(3,1)

$$\max \begin{cases} \text{Down} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{up} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{left} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{Right} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \end{cases}$$

$= -0.04$

(3,2)

$$\max \begin{cases} \text{Down} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{up} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{left} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \\ \text{Right} & -0.04 + 0.95\left((0.7)(0) + \left(\frac{1-0.7}{2}\right)(0) + \left(\frac{1-0.7}{2}\right)(0)\right) \end{cases}$$

$= -0.04$

∴ By Iteration 1

$$\begin{bmatrix} -0.04 & -1 & 1 \\ -0.04 & -0.04 & 0.625 \\ -0.04 & \text{wall} & +0.04 \\ -0.04 & -0.04 & -0.04 \end{bmatrix}$$

	0	1	2
0	-0.04	-1	1
1	-0.04	-0.04	0.625
2	-0.04	Wall	-0.04
3	-0.04	-0.04	-0.04

2nd Iteration

(0,0)

Down: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-1)\right)$

up: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-1)\right)$

left: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

max Right: $-0.04 + 0.95\left((0.7)(-1) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

$= -0.078$

(0,1) -1 : Penalty

(0,2) 1 : Reward.

(1,0)

Down: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

up: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

left: $-0.04 + (0.95)\left[(0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right]$

max Right: $-0.04 + 0.95\left[(0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right]$

$= -0.078$

(1,1)

Down: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(0.625)\right)$

up: $-0.04 + 0.95\left[(0.7)(-1) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(0.625)\right)$

left: $-0.04 + 0.95\left[(0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-1) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

max Right: $-0.04 + 0.95\left[(0.7)(0.625) + \left(\frac{1-0.7}{2}\right)(-1) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

$= 0.2274$

(1,2)

Down: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

up: $-0.04 + 0.95\left((0.7)(1) + \left(\frac{1-0.7}{2}\right)(0.625) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

Left: $-0.04 + 0.95\left((0.7)(0.09) + \left(\frac{1-0.7}{2}\right)(1) + \left(\frac{1-0.7}{2}\right)(0.04)\right)$

max Right: $-0.04 + 0.95\left((0.7)(0.625) + \left(\frac{1-0.7}{2}\right)(1) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

$= 0.70$

(2,0)

Down: $-0.04 + 0.95\left((0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right)$

up: $-0.04 + 0.95\left[(0.7)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04) + \left(\frac{1-0.7}{2}\right)(-0.04)\right]$

left: $-0.04 + 0.95\left[(0.7)(-0.04) + 0.15 \times (-0.04) + 0.15 \times (-0.04)\right]$

max Right: $-0.04 + 0.95\left[(0.7)(-0.04) + 0.15 \times (-0.04) + 0.15 \times (-0.04)\right]$

$= -0.078$

(2,1) 'Wall'

(2,2)

Down: $-0.04 + 0.95\left((0.7)(-0.04) + 0.15 \times (-0.04) + 0.15 \times (-0.04)\right)$

up: $-0.04 + 0.95\left(0.7 \times 0.625 + 0.15(-0.04) + 0.15(-0.04)\right)$

left: $-0.04 + 0.95\left(0.7(-0.04) + 0.15(0.625) + 0.15(0.04)\right)$

max right: $-0.04 + 0.95\left(0.7(-0.04) + 0.15(0.625) + 0.15(-0.04)\right)$

$= 0.36422$

(3,0)

$$\max \begin{cases} \text{Down} & -0.04 + 0.95(0.7(-0.04) + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{up} & -0.04 + 0.95((0.7)(-0.04) + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{left} & -0.04 + 0.95(0.9 \times 0.04 + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{right} & -0.04 + 0.95(0.7 \times 0.04 + 0.15 \times -0.04 + 0.85 \times -0.04) \end{cases}$$

$= -0.078$

(3,1)

$$\max \begin{cases} \text{Down} & -0.04 + 0.95(0.7 \times 0.04 + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{up} & -0.04 + 0.95(0.7 \times 0.04 + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{left} & -0.04 + 0.95(0.7 \times 0.04 + 0.15 \times -0.04 + 0.15 \times 0.04) \\ \text{right} & -0.04 + 0.95(0.7 \times 0.04 + 0.15 \times -0.04 + 0.85 \times -0.04) \end{cases}$$

$= -0.078$

(3,2)

$$\max \begin{cases} \text{Down} & -0.04 + 0.95(0.7 \times 0.04 + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{up} & -0.04 + 0.95(0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{left} & -0.04 + 0.95(0.7 \times -0.04 + 0.15 \times -0.04 + 0.15 \times -0.04) \\ \text{right} & -0.04 + 0.95(0.7 \times 0.04 + 0.15 \times -0.04 + 0.15 \times -0.04) \end{cases}$$

$= -0.078$

∴ After 2nd Iteration

$$\begin{bmatrix} -0.078 & -1 & \\ -0.078 & 0.2274 & 0.70 \\ -0.078 & \text{Wall} & 0.36422 \\ -0.078 & -0.078 & -0.078 \end{bmatrix}$$

Yes values match the output from code.