MDL_ASSIGNMENT 3

Value_iteration

Bellman eq used in algorithm to calculate utility of each grid.

$$U_{t+1}(I) = \max \left[R(I,A) + \gamma P(J/I,A)\, U_t(J) \right]$$

$\gamma \rightarrow$ discount factor i.e $\gamma = 0.95$

$R(I,A) \rightarrow$ Reward taking an action A in state I

$R(I,A) = -0.04$ for ∀ states except reward, penality, wall

algorithm terminates when $|U_t(I) - U_{t+1}(I)| \leq 0.0001$

	1	2	3
1	0	1	-1
2	0	0	0
3	0	0	0
4	0	0	0

→ Initial utility matrix for iteration

Iteration 1

For state : (1,1)

action: ↑

Utility1 = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 1 + 0.15 \times 0)$

= 0.102

action: ↓

utility2 = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 1 + 0.15 \times 0)$

= 0.102

action: →

utility3 = $-0.04 + 0.95(0.7 \times 1 + 0.15 \times 0 + 0.15 \times 0)$

= 0.625

action: ←

utility4 = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 1)$

= 0.094

$U_1(\langle 1,1\rangle)$ = max of utility1, utility2, utility3, utility4

= 0.625

For state (2,1)

action ↑

utility1 = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0)$

= -0.04

action ↓

utility2 = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0)$

= -0.04

action →

utility3 = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0)$

= -0.04

action ←

utility4 = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0)$

= -0.04

$U_1(\langle 2,1\rangle)$ = -0.04

For state (2,2)

utility '↑' = $-0.04 + 0.95(0.7 \times 1 + 0.15 \times 0 + 0.15 \times 0)$

= 0.625

utility ↓ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0)$

= -0.04

utility → = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 1 + 0.15 \times 0)$

= 0.102

utility ← = $0.04 + 0.95(0.7 \times 0 + 0.15 \times 1 + 0.15 \times 0)$

= 0.102

$U_1(\langle 2,2\rangle)$ = 0.625

For state (2,3)

Utility ↑ = $-0.04 + 0.95(0.7 \times (-1) + 0.15 \times 0 + 0.15 \times 0) = -0.7$

Utility ↓ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility → = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times (-1) + 0.15 \times 0) = -0.182$

Utility ← = $-0.04 + 0.95(0.7 \times 0 + 0.15 (-1) + 0.15 \times 0) = -0.182$

$U_1((2,3)) = -0.04$

For state (3,1)

Utility ↑ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility ↓ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility → = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility ← = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

$U_1((3,1)) = -0.04$

For state (3,3)

Utility ↑ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility ↓ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility → = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility ← = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

$U_1((3,3)) = -0.04$

For state (4,1)

Utility ↑ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility ↓ = $-0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

Utility → = $-0.04 + 0.95(0 + 0 + 0) = -0.04$

Utility ← = $-0.04 + 0.95(0 + 0 + 0) = -0.04$

$U_1((4,1)) = -0.04$

For state (4,2)

Utility ↑ = -0.04 + 0.95(0.7 x 0 + 0.15 x 0 + 0.15 x 0) = -0.04

Utility ↓ = -0.04 + 0.95 (0 + 0 + 0) = -0.04

Utility → = -0.04 + 0.95 (0 + 0 + 0) = -0.04

Utility ← = -0.04 + 0.95 (0 + 0 + 0) = -0.04

So $U_1((4,2)) = -0.04$

For state (4,3)

Utility ↑ = -0.04 + 0.95 (0 + 0 + 0) = -0.04

Utility ↓ = -0.04 + 0.95 (0 + 0 + 0) = -0.04

Utility → = -0.04 + 0.95 (0 + 0 + 0) = -0.04

Utility ← = -0.04 + 0.95 (0 + 0 + 0) = -0.04

So $U_2((4,3)) = -0.04$

So Utility matrix after 1st Iteration:

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

Iteration 2:

For state (1,1)

Utility ↑ = -0.04 + 0.95 (0.7 x 0.625 + 0.15 x 0.625 + 0.15 x 1) = 0.607

Utility ↓ = -0.04 + 0.95 (0.7 x (-0.04) + 0.15 x 0.625 + 0.15 x 1) = 0.165

Utility → = -0.04 + 0.95 (0.7 x 1 + 0.15 x 0.625 + 0.15 x -0.04) = 0.708

Utility ← = −0.04 + 0.95 (0.7 × 0.625 + 0.15 × 0.625 + 0.15 × (−0.04)) = 0.459

So $U_2(\langle 1,1 \rangle) = 0.708$

For state (2,1)

Utility ↑ = −0.04 + 0.95 (0.7 × 0.625 + 0.15 × (−0.04) + 0.15 × 0.625) = 0.459

Utility ↓ = −0.04 + 0.95 (0.7 (−0.04) + 0.15 (−0.04) + 0.15 × 0.625) = 0.017

Utility → = −0.04 + 0.95 (0.7 × 0.625 + 0.15 × 0.625 + 0.15 × (−0.04)) = 0.459

Utility ← = −0.04 + 0.95 (0.7 × (−0.04) + 0.15 × 0.625 + 0.15 × (−0.04)) = 0.017

So $U_2(\langle 2,1 \rangle) = 0.459$

For state (2,2)

Utility ↑ = −0.04 + 0.95 (0.7 × 1 + 0.15 × (−0.04) + 0.15 × (−0.04) = **0.614**

Utility ↓ = −0.04 + 0.95 (0.7 × 0.625 + 0.15 (−0.04) + 0.15 × (−0.04) = **0.364**

Utility → = −0.04 + 0.95 (0.7 × (−0.04) + 0.15 (0.625) + 0.15 × (1)) = **0.165**

Utility ← = −0.04 + 0.95 (0.7 × (−0.04) + 0.15 × 1 + 0.15 × 0.625) = **0.165**

So $U_2(\langle 2,2 \rangle) = 0.614$

For state (2,3)

Utility ↑ = $-0.04 + 0.95(0.7 \times (-1) + 0.15 \times 0.625 + 0.15 \times (-0.04) = -0.622$

Utility ↓ = $-0.04 + 0.95(0.7 \times (-0.04) + 0.15 \times 0.625 + 0.15 \times (-0.04) = 0.017$

Utility → = $-0.04 + 0.95(0.7 \times (-0.04) + 0.15 \times (-1) + 0.15 \times (-0.04) = -0.215$

Utility ↓ = $-0.04 + 0.95(0.7 \times 0.625 + 0.15 \times (-1) + 0.15 \times (-0.04) = -0.227$

$U_2(\langle 2,3 \rangle) = 0.227$

For state (3,1)

Utility ↑ = $-0.04 + 0.95(0.7 \times (-0.04) + 0.15 \times (-0.04) + 0.15 \times (-0.04)) = -0.078$

Utility ↓ = $-0.04 + 0.95(0.7 \times (-0.04) + 0.15 \times (-0.04) + 0.15 \times (-0.04)) = -0.078$

Utility → = $-0.04 + 0.95(0.7 \times (-0.04) + 0.15 \times (-0.04) + 0.15 \times (-0.04)) = -0.078$

Utility ← = $-0.04 + 0.95(0.7 \times (-0.04) + 0.15 \times (-0.04) + 0.15 \times (-0.04) = -0.078$

$U_2(\langle 3,1 \rangle) = -0.078$

For State (3,3)

Utility ↑ = $-0.04 + 0.95((0.7 + 0.15 + 0.15)(-0.04)) = -0.078$

Utility ↓ = $-0.04 + 0.95(1 \times (-0.04)) = -0.078$

Utility → = $-0.04 + 0.95(1 \times (-0.04)) = -0.078$

Utility ← = $-0.04 + 0.95(1 \times (-0.04)) = -0.078$

So $U_2(\langle 3,3\rangle) = -0.078$

For State (4,1)

Utility ↑ = $-0.04 + 0.95(1 \times +0.04)) = -0.078$

Utility ↓ = $-0.04 + 0.95(-0.04) = -0.078$

Utility → = $-0.04 + 0.95(-0.04) = -0.078$

Utility ← = $-0.04 + 0.95(-0.04) = -0.078$

So $U_2(\langle 4,1\rangle) = -0.078$

For State (4,2)

Utility ↑ = $-0.04 + 0.95(-0.04) = -0.078$

Utility ↓ = $-0.04 + 0.95(-0.04) = -0.078$

Utility → = $-0.04 + 0.95(-0.04) = -0.078$

Utility ← = $-0.04 + 0.95(-0.04) = -0.078$

So $U_2(\langle 4,2\rangle) = -0.078$

For State (4,3)

Utility ↑ = $-0.04 + 0.95(-0.04) = -0.078$

Utility ↓ = $-0.04 + 0.95(-0.04) = -0.078$

Utility → = $-0.04 + 0.95(0.04) = -0.078$

Utility ← = $-0.04 + 0.95(-0.04) = -0.078$

So $U_2(\langle 4,3\rangle) = -0.078$

So final matrix after 2 iteration :-

0.708	1	-1
0.459	0.614	0.227
-0.078	0	-0.078
-0.078	-0.078	-0.078