9.4 Policy and value iteration

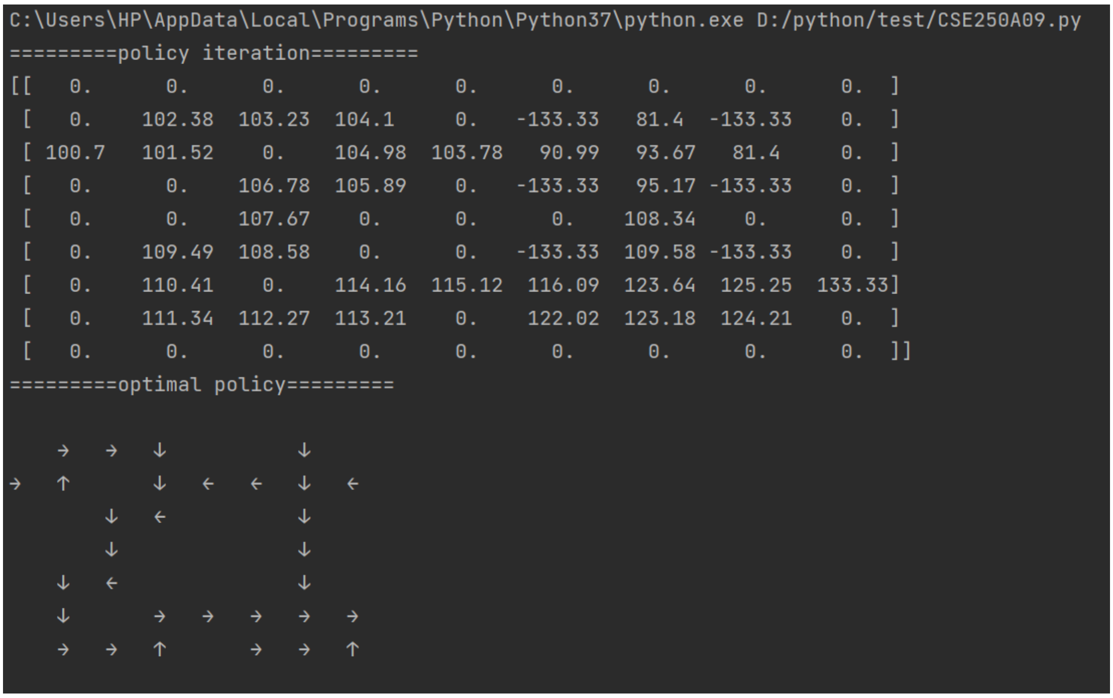
$$V^{\pi}(s) = R(s) + \gamma \sum_{s'} P(s'|s, \pi(s)) V^{\pi}(s')$$

Or

$$R = [I - \gamma * P^{\pi}] V^{\pi}$$

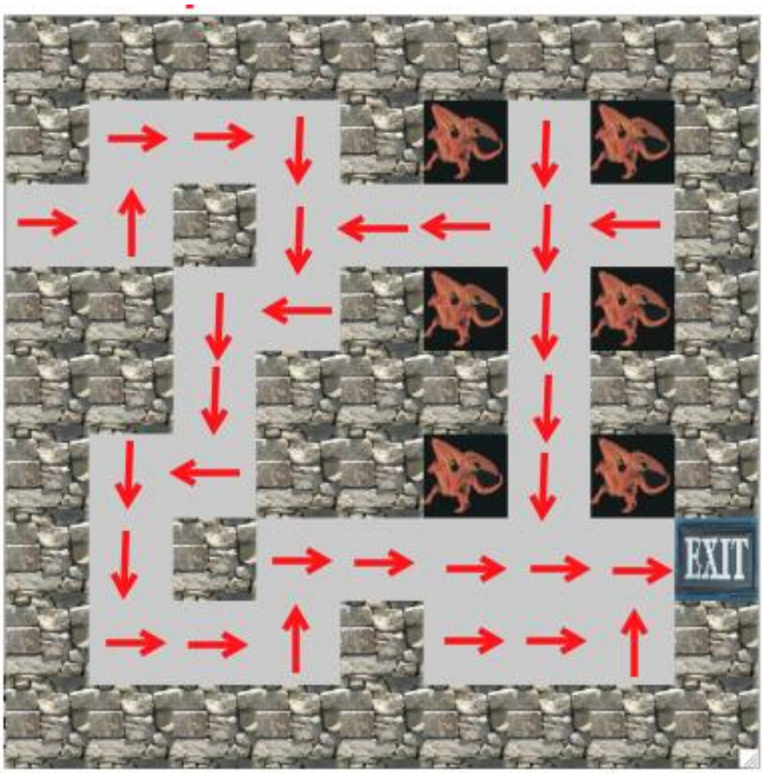
(a)

(i) & (ii)

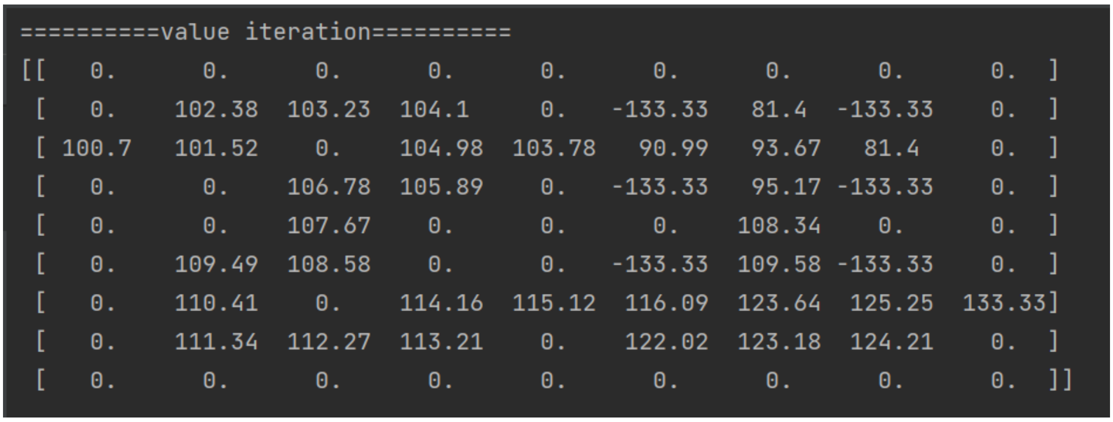


```
C:\Users\HP\AppData\Local\Programs\Python\Python37\python.exe D:/python/test/CSE250A09.py
=========policy iteration=========
[[   0.      0.      0.      0.      0.      0.      0.      0.      0.  ]
 [   0.    102.38  103.23  104.1     0.   -133.33   81.4  -133.33    0.  ]
 [ 100.7   101.52    0.    104.98  103.78   90.99   93.67   81.4     0.  ]
 [   0.      0.    106.78  105.89    0.   -133.33   95.17 -133.33    0.  ]
 [   0.      0.    107.67    0.      0.      0.    108.34    0.      0.  ]
 [   0.    109.49  108.58    0.      0.   -133.33  109.58 -133.33    0.  ]
 [   0.    110.41    0.    114.16  115.12  116.09  123.64  125.25  133.33]
 [   0.    111.34  112.27  113.21    0.    122.02  123.18  124.21    0.  ]
 [   0.      0.      0.      0.      0.      0.      0.      0.      0.  ]]
=========optimal policy=========

    →   →   ↓           ↓
→   ↑       ↓   ←   ←   ↓   ←
        ↓   ←           ↓
        ↓               ↓
    ↓   ←               ↓
    ↓       →   →   →   →   →
    →   →   ↑       →   →   ↑

```

	102.38	103.23	104.10		-133.33	81.4	-133.33	
100.70	101.52		104.98	103.78	90.99	93.67	81.4	
		106.78	105.89		-133.33	95.17	-133.33	
		107.67				108.34		
	109.49	108.58			-133.33	109.58	-133.33	
	110.41		114.16	115.12	116.09	123.64	125.25	133.33
	111.34	112.27	113.21		122.02	123.18	124.21	



(b)



```
==========value iteration==========
[[   0.      0.      0.      0.      0.      0.      0.      0.      0.  ]
 [   0.    102.38  103.23  104.1     0.   -133.33   81.4  -133.33    0.  ]
 [ 100.7   101.52    0.    104.98  103.78   90.99   93.67   81.4     0.  ]
 [   0.      0.    106.78  105.89    0.   -133.33   95.17 -133.33    0.  ]
 [   0.      0.    107.67    0.      0.      0.    108.34    0.      0.  ]
 [   0.    109.49  108.58    0.      0.   -133.33  109.58 -133.33    0.  ]
 [   0.    110.41    0.    114.16  115.12  116.09  123.64  125.25  133.33]
 [   0.    111.34  112.27  113.21    0.    122.02  123.18  124.21    0.  ]
 [   0.      0.      0.      0.      0.      0.      0.      0.      0.  ]]
```

Agree with the result from part(a)

	102.38	103.23	104.10		-133.33	81.4	-133.33	
100.70	101.52		104.98	103.78	90.99	93.67	81.4	
		106.78	105.89		-133.33	95.17	-133.33	
		107.67				108.34		
	109.49	108.58			-133.33	109.58	-133.33	
	110.41		114.16	115.12	116.09	123.64	125.25	133.33
	111.34	112.27	113.21		122.02	123.18	124.21	