PROBLEM 3
PART-2

MARKOV DECISION PROBLEM (MDP)
POLICY ITERATION

Problem 2: Policy Iteration for both version of Magneto on MDP

Maximum expectation possible: +20
Minimum expectation possible: -20

On the basis of the random moves of both jean and magneto(case 1), the expectation table will be created. [Each output test case is run for 200 iterations only if does not converge before 200 iterations]

Note: Coordinates of the board are considered as given below (x_cordinate,y_cordinate):

(0, 0) |(0, 1) |(0, 2) |(0, 3) |(0, 4) |
(1, 0) |(1, 1) |(1, 2) |(1, 3) |(1, 4) |
(2, 0) |(2, 1) |(2, 2) |(2, 3) |(2, 4) |
(3, 0) |(3, 1) |(3, 2) |(3, 3) |(3, 4) |
(4, 0) |(4, 1) |(4, 2) |(4, 3) |(4, 4) |

U,D,L,R,X stands for move up,down,left,right and no move in the board output.

Output: [Lazy Magneto ----Sol_2_PI_Lazy.ipynb]

Initial rewards configurations:

```
-----------------------------------
 0.00| 0.00| 0.00| 0.00| 0.00|
-----------------------------------
 0.00| 0.00| 0.00| 0.00| 0.00|
-----------------------------------
 0.00| 0.00| 0.00| 0.00| 0.00|
-----------------------------------
 0.00| 0.00| 0.00| 0.00|-20.00|
-----------------------------------
 0.00| 20.00| 0.00| 0.00| 0.00|
```

Initial random policy

R	R	U	D	L
D	R	D	D	R
L	D	D	L	D
U	D		U	D
D	X	X	D	R

Initial random probability

0.57	1.00	0.64	0.87	0.52
0.94	0.15	0.71	0.35	0.65
0.21	0.29	0.93	0.17	0.63
0.74	0.27	0.00	0.41	0.04
0.62	0.68	0.85	0.08	0.08

After 200 Policy iterations:

Values 200:

```
----------------------------------
-5.00|-0.62| 10.00| 20.00| 10.00|
----------------------------------
-10.00|-0.31|-0.62| 10.00|-0.62|
----------------------------------
-10.00|-0.62|-20.00|-10.00|-1.25|
----------------------------------
-10.00| 10.00| 0.00|-0.00|-0.00|
----------------------------------
 2.50|-10.00|-20.00|-0.00|-0.00|
```

Policy 200:

```
-------------------------------
  D  |  R  |  R  |  X  |  L  |
-------------------------------
  U  |  X  |  U  |  U  |  L  |
-------------------------------
  R  |  D  |  U  |  U  |  U  |
-------------------------------
  R  |  D  |     |  D  |  D  |
-------------------------------
  R  |  X  |  L  |  R  |  U  |
```

FOR ACTIVE MAGNETO MOVE OUTPUTS:

Output: [Active Magneto---- Sol_2_Pl_Active.ipynb]

Initial rewards configurations:

```
----------------------------------
-20.00| 0.00| 0.00| 0.00| 0.00|
----------------------------------
 0.00| 0.00| 0.00| 0.00| 0.00|
----------------------------------
 0.00| 0.00| 0.00| 0.00| 0.00|
----------------------------------
 0.00| 0.00| 0.00| 0.00| 0.00|
----------------------------------
 0.00| 20.00| 0.00| 0.00| 0.00|
```

Initial random policy

```
-------------------------------
 D  |  X  |  L  |  D  |  L  |
-------------------------------
 X  |  L  |  R  |  R  |  U  |
-------------------------------
 U  |  X  |  L  |  U  |  D  |
-------------------------------
 R  |  R  |     |  X  |  R  |
-------------------------------
 U  |  R  |  U  |  R  |  D  |
```

Initial random probability

```
-----------------------------------
 0.97| 0.06| 0.78| 0.44| 0.83|
-----------------------------------
 0.43| 0.30| 0.62| 0.57| 0.52|
-----------------------------------
 0.60| 0.00| 0.14| 0.89| 0.68|
-----------------------------------
 0.20| 0.11| 0.00| 0.63| 0.16|
-----------------------------------
 0.51| 0.73| 0.92| 0.32| 0.70|
```

After 200 Policy iterations:

Values 200:

```
-----------------------------------
 10.44| 12.28| 10.44|-20.00| 17.00|
-----------------------------------
```

 12.28| 10.44| 12.28| 10.44| 20.00|

 12.28| 12.28| 14.45| 12.28|-0.13|

 14.45| 14.45| 0.00| 10.44|-0.13|

 14.45| 20.00| 0.00| 7.54| 6.41|

Policy 200:

D	L	D	R	D
U	U	D	L	X
D	D	X	L	U
D	D		U	D
R	X	L	U	L