1. Value iteration results

```
$ python3 mdpVI.py
16:30:24
----------------------------------------------------
output for value iteration with params (-2, 0.9)
State: (1, 1), val: -3.0514926010225425, policy: up
State: (1, 2), val: -0.8078741742648561, policy: up
State: (1, 3), val: 1.8577098725883507, policy: right
State: (2, 1), val: -2.1686795408094994, policy: right
State: (2, 3), val: 5.2267109404105, policy: right
State: (3, 1), val: 0.3079009060173786, policy: up
State: (3, 2), val: 3.830672847532406, policy: up
State: (3, 3), val: 8.730431946440158, policy: right
State: (4, 1), val: -2.8333004284273655, policy: left
State: (4, 2), val: -9.999916647515823, policy: NA
State: (4, 3), val: 9.999916647515823, policy: NA
----------------------------------------------------
----------------------------------------------------
output for value iteration with params (-0.2, 0.9)
State: (1, 1), val: 4.587181682454209, policy: up
State: (1, 2), val: 5.5822857276050275, policy: up
State: (1, 3), val: 6.635391983051275, policy: right
State: (2, 1), val: 3.945525666051915, policy: right
State: (2, 3), val: 7.96640028673407, policy: right
State: (3, 1), val: 4.771301280498433, policy: up
State: (3, 2), val: 6.1016017090547425, policy: up
State: (3, 3), val: 9.350633701975335, policy: right
State: (4, 1), val: 2.4784009667664355, policy: left
State: (4, 2), val: -9.999916647515823, policy: NA
State: (4, 3), val: 9.999916647515823, policy: NA
----------------------------------------------------
----------------------------------------------------
output for value iteration with params (-0.01, 0.9)
State: (1, 1), val: 5.407851054450954, policy: up
State: (1, 2), val: 6.256802606135739, policy: up
State: (1, 3), val: 7.139702872489028, policy: right
State: (2, 1), val: 4.736160181976276, policy: left
State: (2, 3), val: 8.25558971773489, policy: right
State: (3, 1), val: 5.256512352112594, policy: up
State: (3, 2), val: 6.341310866659878, policy: up
State: (3, 3), val: 9.416099442837382, policy: right
State: (4, 1), val: 3.050216310033905, policy: left
State: (4, 2), val: -9.999916647515823, policy: NA
State: (4, 3), val: 9.999916647515823, policy: NA
```

2. Output for policy iteration

```
$ python3 mdpPI.py
```

```
16:32:14
--------------------------------------------------
output for policy iteration with params (-2, 0.9)
State: (1, 1), val: -3.051416476886305, policy: up
State: (1, 2), val: -0.8077952729827622, policy: up
State: (1, 3), val: 1.8577887752322593, policy: right
State: (2, 1), val: -2.168621085671812, policy: right
State: (2, 3), val: 5.226790829549925, policy: right
State: (3, 1), val: 0.30795936115506584, policy: up
State: (3, 2), val: 3.8307346464447587, policy: up
State: (3, 3), val: 8.730511835579584, policy: right
State: (4, 1), val: -2.8332628779732394, policy: left
State: (4, 2), val: -9.999999570420037, policy: NO POLICY For TERM
States
State: (4, 3), val: 9.999999570420037, policy: NO POLICY For TERM
States
--------------------------------------------------
--------------------------------------------------
output for policy iteration with params (-0.2, 0.9)
State: (1, 1), val: 4.58725852995975, policy: up
State: (1, 2), val: 5.582365039212962, policy: up
State: (1, 3), val: 6.635471294659208, policy: right
State: (2, 1), val: 3.9455844240149203, policy: right
State: (2, 3), val: 7.9664805897371025, policy: right
State: (3, 1), val: 4.771360038461438, policy: up
State: (3, 2), val: 6.101663828114754, policy: up
State: (3, 3), val: 9.350714004978368, policy: right
State: (4, 1), val: 2.4784387117497113, policy: left
State: (4, 2), val: -9.999999999999984, policy: NO POLICY For TERM
States
State: (4, 3), val: 9.999999999999984, policy: NO POLICY For TERM
States
--------------------------------------------------
--------------------------------------------------
output for policy iteration with params (-0.01, 0.9)
State: (1, 1), val: 5.40792880644135, policy: up
State: (1, 2), val: 6.25688191771565, policy: up
State: (1, 3), val: 7.139782184068953, policy: right
State: (2, 1), val: 4.736229480738858, policy: left
State: (2, 3), val: 8.255670020709568, policy: right
State: (3, 1), val: 5.256572093129815, policy: up
State: (3, 2), val: 6.341372985697955, policy: up
State: (3, 3), val: 9.416179745812057, policy: right
State: (4, 1), val: 3.05025479721896, policy: left
State: (4, 2), val: -9.99999999997055, policy: NO POLICY For TERM
States
State: (4, 3), val: 9.99999999997055, policy: NO POLICY For TERM
States
```

3. Output for monte carlo on policy first visit epislon soft policy

```
$ python3 mdpMC.py
16:32:54
--------------------------------------------------
output for monte carlo with params (-2, 0.9, and 0.1)
state:             action-values q(s,a),           policy
(1, 1): up=-6.8452527512436, down=0, left=-13.077160244514248,
right=-6.550981674781947  policy=down
(1, 2): up=-5.18032612270257, down=-14.5227064430829,
left=-6.959996993952795, right=-7.168661341268301  policy=up
(1, 3): up=-5.336868566578569, down=-7.187592902745025,
left=-5.282519239354788, right=-3.372633146385928  policy=right
(2, 1): up=-6.743018380460459, down=-7.315222523117705,
left=-14.030180389481373, right=-4.872731350115201  policy=right
(2, 3): up=-3.3855373208000046, down=-3.436221424460436,
left=-5.132068516129036, right=-1.344545725959725  policy=right
(3, 1): up=-3.1892947631136295, down=-4.895085708029206,
left=-6.746474835841002, right=-3.0670275747192015  policy=right
(3, 2): up=-1.1, down=-6.612580873100001, left=-4.033014506923076,
right=-1.0  policy=right
(3, 3): up=-1.2064117647058796, down=-3.033126666666661,
left=-3.323777857142862, right=1.0  policy=right
(4, 1): up=-1.0, down=-3.193468571428569, left=-4.947646706428581,
right=-3.038954330708656  policy=up
--------------------------------------------------
--------------------------------------------------
output for monte carlo with params (-0.2, 0.9, and 0.1)
state:             action-values q(s,a),           policy
(1, 1): up=0, down=-1.9800458080290628, left=-1.9531938706854572,
right=-0.7221858063958141  policy=up
(1, 2): up=0.1320419854498389, down=-1.1480880952493975,
left=-0.031700000000000006, right=-1.0406001884335057  policy=up
(1, 3): up=0.060760453634756545, down=-0.5268584016365767,
left=0.1253592043531643, right=0.42262208884097247  policy=right
(2, 1): up=-1.6151302145245312, down=-1.7850673687930876,
left=-1.9640824544523139, right=-0.5896954260875042  policy=right
(2, 3): up=-0.07661715807280187, down=0.35189589584714154,
left=0.15555927445946002, right=0.670068244148257  policy=right
(3, 1): up=0.22737887339303892, down=-1.9642224046760948,
left=-1.6795697415130704, right=-1.567193007316976  policy=up
(3, 2): up=0.6673901242603496, down=-0.7323404895588649,
left=0.052972821999999954, right=-1.0  policy=up
(3, 3): up=0.6654540963855411, down=0.3471352692307706,
left=0.3647178907000017, right=1.0  policy=right
(4, 1): up=-1.0, down=-1.6568225466884219, left=-1.5147022859720458,
right=-1.1953625322005241  policy=up
--------------------------------------------------
--------------------------------------------------
output for monte carlo with params (-0.01, 0.9, and 0.1)
```

```
state:             action-values q(s,a),           policy
(1, 1): up=0.5973686900000001, down=0, left=0,
right=-0.6314410000000001  policy=up
(1, 2): up=0.6981726216061227, down=-0.022941054999999988,
left=0.6217100000000001, right=0.6217100000000001  policy=up
(1, 3): up=0.7002293749999994, down=0.6217100000000001,
left=0.7005174137931033, right=0.7879424599265422  policy=right
(2, 1): up=-0.7232950000000001, down=0, left=0.5254411031000001,
right=-0.8290000000000001  policy=left
(2, 3): up=0.7799685714285687, down=0.7789615261194004,
left=0.6956330504201654, right=0.8775384796066918  policy=right
(3, 1): up=0.224, down=0, left=0.4845851000000002, right=0
policy=left
(3, 2): up=0.872985357142852, down=0.51805106, left=0.791, right=-1.0
policy=up
(3, 3): up=0.8800999999999942, down=0.7152553423076901,
left=0.7723996068965492, right=1.0  policy=right
(4, 1): up=0, down=0, left=0, right=0  policy=up
--------------------------------------------------
```