output

---------- Part 1 Montecarlo First Visit ----------

Epoch 0
N(s):
[[0. 1. 0. 0. 0.]
[1. 1. 0. 0. 0.]
[1. 0. 0. 0. 0.]
[1. 0. 0. 0. 0.]
[1. 0. 0. 0. 0.]]
S(s):
[[0. -1.9 0. 0. 0.]
[-3.439 -2.71 0. 0. 0.]
[-4.0951 0. 0. 0. 0.]
[-4.68559 0. 0. 0. 0.]
[-5.217031 0. 0. 0. 0.]]
V(s):
[[0. -1.9 0. 0. 0.]
[-3.439 -2.71 0. 0. 0.]
[-4.0951 0. 0. 0. 0.]
[-4.68559 0. 0. 0. 0.]
[-5.217031 0. 0. 0. 0.]]

Returns for this episode:
k s r Y G(s)
1 (4, 0) -1 0.9 -5.217
2 (3, 0) -1 0.9 -4.686
3 (2, 0) -1 0.9 -4.095
4 (1, 0) -1 0.9 -3.439
5 (1, 1) -1 0.9 -2.710
6 (0, 1) -1 0.9 -1.900
7 (1, 1) -1 0.9 -1.000
8 (1, 0) 0 0.9 0.000

Epoch 1
N(s):
[[0. 1. 1. 1. 1.]
[2. 2. 1. 1. 1.]

[2. 1. 1. 1. 1.]
[2. 1. 1. 1. 1.]
[2. 1. 1. 1. 0.]]
S(s):
[[0. -1.9 -9.99757251 -9.99781525 -9.99870993]
[-13.42444422 -12.69382691 -9.99730278 -9.99895504 -9.99883894]
[-14.08199979 -9.98820982 -9.99944467 -9.99938296 -9.99840732]
[-14.67785446 -9.99140496 -9.99588902 -9.9993144 -3.439]
[-15.21139179 -9.99492471 -9.99543224 -9.99923823 0.]]
V(s):
[[0. -1.9 -9.99757251 -9.99781525 -9.99870993]
[-6.71222211 -6.34691345 -9.99730278 -9.99895504 -9.99883894]
[-7.0409999 -9.98820982 -9.99944467 -9.99938296 -9.99840732]
[-7.33892723 -9.99140496 -9.99588902 -9.9993144 -3.439]
[-7.6056959 -9.99492471 -9.99543224 -9.99923823 0.]]

Returns for this episode:
k s r Y G(s)
1 (2, 2) -1 0.9 -9.999
2 (2, 3) -1 0.9 -9.999
3 (3, 3) -1 0.9 -9.999
4 (4, 3) -1 0.9 -9.999
5 (3, 3) -1 0.9 -9.999
6 (2, 3) -1 0.9 -9.999
7 (1, 3) -1 0.9 -9.999
8 (1, 4) -1 0.9 -9.999
9 (0, 4) -1 0.9 -9.999
10 (1, 4) -1 0.9 -9.999
11 (2, 4) -1 0.9 -9.998
12 (1, 4) -1 0.9 -9.998
13 (0, 4) -1 0.9 -9.998
14 (0, 3) -1 0.9 -9.998
15 (0, 2) -1 0.9 -9.998
16 (1, 2) -1 0.9 -9.997
17 (2, 2) -1 0.9 -9.997
18 (1, 2) -1 0.9 -9.997
19 (2, 2) -1 0.9 -9.996
20 (3, 2) -1 0.9 -9.996
21 (4, 2) -1 0.9 -9.995
22 (4, 1) -1 0.9 -9.995
23 (4, 0) -1 0.9 -9.994

24 (4, 1) -1 0.9 -9.994
25 (4, 0) -1 0.9 -9.993
26 (3, 0) -1 0.9 -9.992
27 (3, 1) -1 0.9 -9.991
28 (4, 1) -1 0.9 -9.990
29 (3, 1) -1 0.9 -9.989
30 (2, 1) -1 0.9 -9.988
31 (2, 0) -1 0.9 -9.987
32 (1, 0) -1 0.9 -9.985
33 (1, 1) -1 0.9 -9.984
34 (1, 0) -1 0.9 -9.982
35 (2, 0) -1 0.9 -9.980
36 (2, 1) -1 0.9 -9.978
37 (2, 0) -1 0.9 -9.975
38 (2, 1) -1 0.9 -9.973
39 (2, 0) -1 0.9 -9.970
40 (2, 1) -1 0.9 -9.966
41 (2, 0) -1 0.9 -9.962
42 (2, 1) -1 0.9 -9.958
43 (2, 2) -1 0.9 -9.954
44 (1, 2) -1 0.9 -9.948
45 (1, 3) -1 0.9 -9.943
46 (1, 4) -1 0.9 -9.936
47 (1, 3) -1 0.9 -9.929
48 (0, 3) -1 0.9 -9.921
49 (0, 4) -1 0.9 -9.913
50 (0, 3) -1 0.9 -9.903
51 (1, 3) -1 0.9 -9.892
52 (2, 3) -1 0.9 -9.880
53 (1, 3) -1 0.9 -9.867
54 (1, 4) -1 0.9 -9.852
55 (2, 4) -1 0.9 -9.836
56 (2, 3) -1 0.9 -9.818
57 (3, 3) -1 0.9 -9.797
58 (4, 3) -1 0.9 -9.775
59 (3, 3) -1 0.9 -9.750
60 (2, 3) -1 0.9 -9.722
61 (1, 3) -1 0.9 -9.691
62 (1, 4) -1 0.9 -9.657
63 (2, 4) -1 0.9 -9.618

64 (2, 3) -1 0.9 -9.576
65 (2, 4) -1 0.9 -9.529
66 (1, 4) -1 0.9 -9.477
67 (0, 4) -1 0.9 -9.419
68 (0, 3) -1 0.9 -9.354
69 (0, 4) -1 0.9 -9.282
70 (1, 4) -1 0.9 -9.202
71 (0, 4) -1 0.9 -9.114
72 (0, 3) -1 0.9 -9.015
73 (0, 2) -1 0.9 -8.906
74 (0, 3) -1 0.9 -8.784
75 (0, 2) -1 0.9 -8.649
76 (1, 2) -1 0.9 -8.499
77 (2, 2) -1 0.9 -8.332
78 (2, 3) -1 0.9 -8.147
79 (1, 3) -1 0.9 -7.941
80 (1, 2) -1 0.9 -7.712
81 (0, 2) -1 0.9 -7.458
82 (0, 3) -1 0.9 -7.176
83 (0, 4) -1 0.9 -6.862
84 (0, 3) -1 0.9 -6.513
85 (0, 2) -1 0.9 -6.126
86 (0, 3) -1 0.9 -5.695
87 (1, 3) -1 0.9 -5.217
88 (2, 3) -1 0.9 -4.686
89 (3, 3) -1 0.9 -4.095
90 (3, 4) -1 0.9 -3.439
91 (3, 3) -1 0.9 -2.710
92 (3, 2) -1 0.9 -1.900
93 (4, 2) -1 0.9 -1.000
94 (4, 3) 0 0.9 0.000

Epoch 10
N(s):
[[0. 3. 4. 4. 5.]
[6. 5. 5. 5. 5.]
[7. 6. 6. 7. 5.]
[6. 6. 6. 8. 3.]
[6. 4. 5. 8. 0.]]
S(s):
[[0. -5.339 -26.69491987 -36.95722593 -46.51154017]

[-39.49196477 -35.38093372 -41.75214075 -46.05785551 -40.47666223]
[-59.08929248 -56.01809137 -54.51087047 -53.45247505 -38.03693367]
[-52.12958579 -56.82500215 -46.1635894 -52.12291461 -13.05132075]
[-49.41734639 -36.34204112 -35.21636538 -33.1420852 0.]]
V(s):
[[0. -1.77966667 -6.67372997 -9.23930648 -9.30230803]
[-6.58199413 -7.07618674 -8.35042815 -9.2115711 -8.09533245]
[-8.4413275 -9.33634856 -9.08514508 -7.63606786 -7.60738673]
[-8.6882643 -9.47083369 -7.69393157 -6.51536433 -4.35044025]
[-8.2362244 -9.08551028 -7.04327308 -4.14276065 0.]]

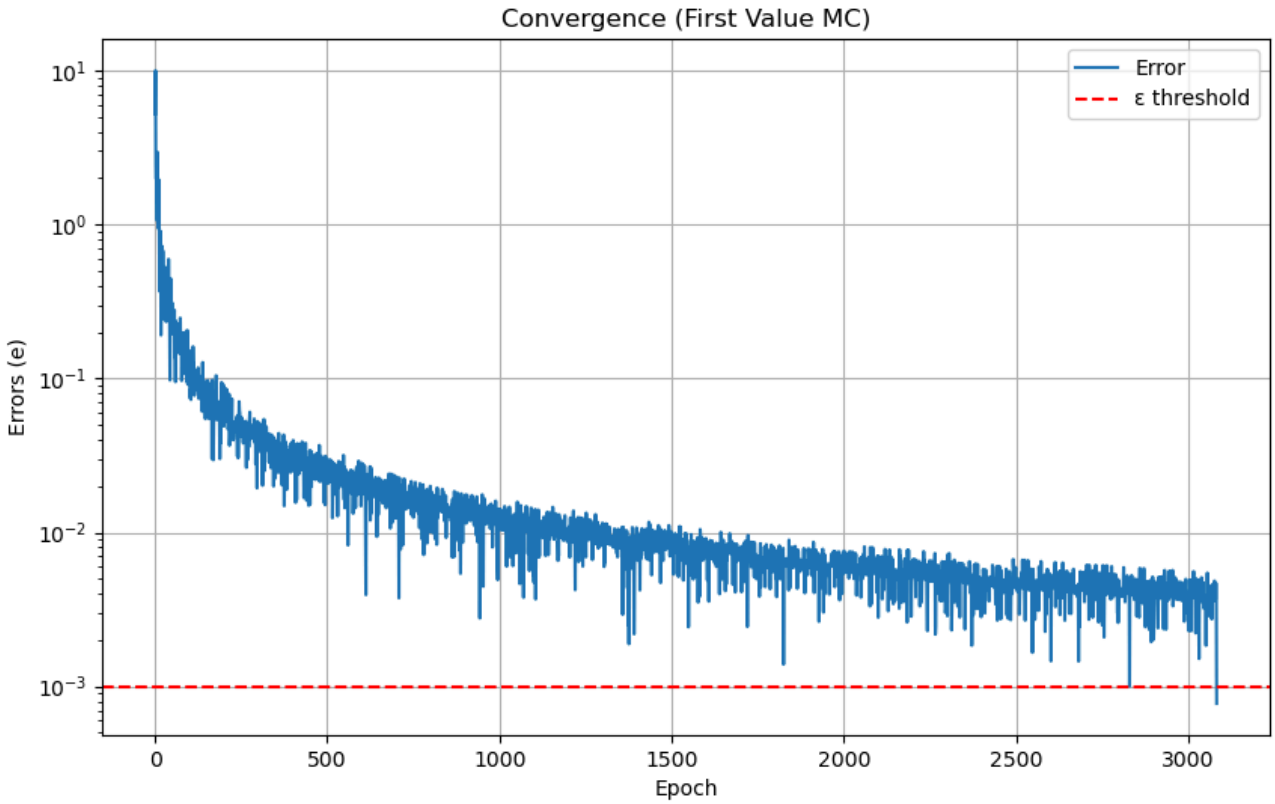
Returns for this episode:
k s r Y G(s)
1 (0, 4) -1 0.9 -9.975
2 (1, 4) -1 0.9 -9.973
3 (0, 4) -1 0.9 -9.970
4 (0, 3) -1 0.9 -9.966
5 (0, 4) -1 0.9 -9.962
6 (0, 3) -1 0.9 -9.958
7 (0, 4) -1 0.9 -9.954
8 (1, 4) -1 0.9 -9.948
9 (1, 3) -1 0.9 -9.943
10 (1, 4) -1 0.9 -9.936
11 (2, 4) -1 0.9 -9.929
12 (2, 3) -1 0.9 -9.921
13 (3, 3) -1 0.9 -9.913
14 (3, 2) -1 0.9 -9.903
15 (2, 2) -1 0.9 -9.892
16 (2, 3) -1 0.9 -9.880
17 (3, 3) -1 0.9 -9.867
18 (4, 3) -1 0.9 -9.852
19 (4, 2) -1 0.9 -9.836
20 (3, 2) -1 0.9 -9.818
21 (2, 2) -1 0.9 -9.797
22 (2, 3) -1 0.9 -9.775
23 (3, 3) -1 0.9 -9.750
24 (4, 3) -1 0.9 -9.722
25 (3, 3) -1 0.9 -9.691
26 (3, 2) -1 0.9 -9.657
27 (2, 2) -1 0.9 -9.618
28 (3, 2) -1 0.9 -9.576

29 (3, 3) -1 0.9 -9.529
30 (3, 2) -1 0.9 -9.477
31 (3, 1) -1 0.9 -9.419
32 (4, 1) -1 0.9 -9.354
33 (4, 0) -1 0.9 -9.282
34 (3, 0) -1 0.9 -9.202
35 (4, 0) -1 0.9 -9.114
36 (3, 0) -1 0.9 -9.015
37 (2, 0) -1 0.9 -8.906
38 (1, 0) -1 0.9 -8.784
39 (2, 0) -1 0.9 -8.649
40 (3, 0) -1 0.9 -8.499
41 (2, 0) -1 0.9 -8.332
42 (1, 0) -1 0.9 -8.147
43 (1, 1) -1 0.9 -7.941
44 (1, 2) -1 0.9 -7.712
45 (1, 3) -1 0.9 -7.458
46 (1, 2) -1 0.9 -7.176
47 (1, 3) -1 0.9 -6.862
48 (0, 3) -1 0.9 -6.513
49 (0, 4) -1 0.9 -6.126
50 (1, 4) -1 0.9 -5.695
51 (0, 4) -1 0.9 -5.217
52 (1, 4) -1 0.9 -4.686
53 (1, 3) -1 0.9 -4.095
54 (1, 2) -1 0.9 -3.439
55 (0, 2) -1 0.9 -2.710
56 (1, 2) -1 0.9 -1.900
57 (0, 2) -1 0.9 -1.000
58 (0, 1) 0 0.9 0.000

Epoch 3083
N(s):
[[0. 1273. 1260. 1215. 1002.]
[1312. 1577. 1575. 1511. 1244.]
[1337. 1583. 1661. 1660. 1330.]
[1256. 1492. 1616. 1624. 1329.]
[1018. 1244. 1358. 1340. 0.]]
S(s):
[[0. -6175.9663643 -8948.00924618 -9759.92956613
-8254.87531538]

[-6237.1050454 -10450.44628002 -12179.63231874 -12180.3844324
-10043.19704585]
[-9514.34613922 -12213.16938384 -13247.77445864 -12813.54552184
-9606.88550619]
[-9959.81611641 -11966.45886807 -12453.53123907 -10893.13656832
-6489.80053883]
[-8341.888477 -9833.90502229 -9534.88137309 -6276.41485053
0.]]
V(s):
[[0. -4.85150539 -7.10159464 -8.03286384 -8.23839852]
[-4.75389104 -6.62678902 -7.73309988 -8.06114125 -8.07330952]
[-7.11619008 -7.71520492 -7.97578234 -7.71900333 -7.22322219]
[-7.9297899 -8.02041479 -7.70639309 -6.70759641 -4.88322087]
[-8.19438947 -7.90506835 -7.02126758 -4.68389168 0.]]

Returns for this episode:
k s r Y G(s)
1 (2, 2) -1 0.9 -8.332
2 (3, 2) -1 0.9 -8.147
3 (4, 2) -1 0.9 -7.941
4 (3, 2) -1 0.9 -7.712
5 (2, 2) -1 0.9 -7.458
6 (2, 1) -1 0.9 -7.176
7 (2, 2) -1 0.9 -6.862
8 (1, 2) -1 0.9 -6.513
9 (0, 2) -1 0.9 -6.126
10 (1, 2) -1 0.9 -5.695
11 (0, 2) -1 0.9 -5.217
12 (0, 1) -1 0.9 -4.686
13 (0, 2) -1 0.9 -4.095
14 (1, 2) -1 0.9 -3.439
15 (0, 2) -1 0.9 -2.710
16 (0, 1) -1 0.9 -1.900
17 (0, 2) -1 0.9 -1.000
18 (0, 1) 0 0.9 0.000



---------- Part 2 Montecarlo Every Visit ----------

Epoch 0
N(s):
[[0. 0. 2. 1. 1.]
[2. 1. 3. 1. 0.]
[3. 3. 1. 2. 0.]
[0. 1. 0. 1. 0.]
[0. 0. 0. 0. 0.]]
S(s):
[[0. 0. -16.98143011 -8.78423345 -8.90581011]
[-5.6953279 -6.12579511 -23.15924906 -6.86189404 0.]
[-8.927031 -10.02459 -7.94108868 -14.88802539 0.]
[0. -4.0951 0. -7.45813417 0.]
[0. 0. 0. 0. 0.]]
V(s):
[[0. 0. -8.49071506 -8.78423345 -8.90581011]
[-2.84766395 -6.12579511 -7.71974969 -6.86189404 0.]
[-2.975677 -3.34153 -7.94108868 -7.44401269 0.]
[0. -4.0951 0. -7.45813417 0.]
[0. 0. 0. 0. 0.]]

Returns for this episode:
k s r Y G(s)
1 (0, 4) -1 0.9 -8.906
2 (0, 3) -1 0.9 -8.784
3 (0, 2) -1 0.9 -8.649
4 (1, 2) -1 0.9 -8.499
5 (0, 2) -1 0.9 -8.332
6 (1, 2) -1 0.9 -8.147
7 (2, 2) -1 0.9 -7.941
8 (2, 3) -1 0.9 -7.712
9 (3, 3) -1 0.9 -7.458
10 (2, 3) -1 0.9 -7.176
11 (1, 3) -1 0.9 -6.862
12 (1, 2) -1 0.9 -6.513
13 (1, 1) -1 0.9 -6.126
14 (1, 0) -1 0.9 -5.695
15 (2, 0) -1 0.9 -5.217
16 (2, 1) -1 0.9 -4.686
17 (3, 1) -1 0.9 -4.095
18 (2, 1) -1 0.9 -3.439
19 (2, 0) -1 0.9 -2.710
20 (2, 1) -1 0.9 -1.900
21 (2, 0) -1 0.9 -1.000
22 (1, 0) 0 0.9 0.000

Epoch 1
N(s):
[[0. 2. 7. 8. 6.]
[2. 1. 6. 4. 2.]
[3. 3. 2. 3. 0.]
[0. 1. 0. 1. 0.]
[0. 0. 0. 0. 0.]]
S(s):
[[0. -4.68559 -43.43755716 -65.08602303 -50.19227318]
[-5.6953279 -6.12579511 -35.01146466 -31.51989797 -17.2832871]
[-8.927031 -10.02459 -10.65108868 -24.36467776 0.]
[0. -4.0951 0. -7.45813417 0.]
[0. 0. 0. 0. 0.]]
V(s):
[[0. -2.342795 -6.20536531 -8.13575288 -8.36537886]
[-2.84766395 -6.12579511 -5.83524411 -7.87997449 -8.64164355]

[-2.975677 -3.34153 -5.32554434 -8.12155925 0.]
[0. -4.0951 0. -7.45813417 0.]
[0. 0. 0. 0. 0.]]

Returns for this episode:
k s r Y G(s)
1 (2, 3) -1 0.9 -9.477
2 (1, 3) -1 0.9 -9.419
3 (0, 3) -1 0.9 -9.354
4 (0, 2) -1 0.9 -9.282
5 (0, 3) -1 0.9 -9.202
6 (1, 3) -1 0.9 -9.114
7 (0, 3) -1 0.9 -9.015
8 (0, 4) -1 0.9 -8.906
9 (1, 4) -1 0.9 -8.784
10 (0, 4) -1 0.9 -8.649
11 (1, 4) -1 0.9 -8.499
12 (0, 4) -1 0.9 -8.332
13 (0, 3) -1 0.9 -8.147
14 (0, 4) -1 0.9 -7.941
15 (0, 3) -1 0.9 -7.712
16 (0, 4) -1 0.9 -7.458
17 (0, 3) -1 0.9 -7.176
18 (0, 2) -1 0.9 -6.862
19 (1, 2) -1 0.9 -6.513
20 (1, 3) -1 0.9 -6.126
21 (0, 3) -1 0.9 -5.695
22 (0, 2) -1 0.9 -5.217
23 (0, 1) -1 0.9 -4.686
24 (0, 2) -1 0.9 -4.095
25 (1, 2) -1 0.9 -3.439
26 (2, 2) -1 0.9 -2.710
27 (1, 2) -1 0.9 -1.900
28 (0, 2) -1 0.9 -1.000
29 (0, 1) 0 0.9 0.000

Epoch 10
N(s):
[[0. 6. 12. 10. 6.]
[5. 13. 21. 15. 10.]
[12. 21. 20. 18. 11.]

[15. 30. 20. 15. 8.]
[10. 27. 16. 5. 0.]]
S(s):
[[0. -18.44384442 -79.00314361 -80.68437356 -50.19227318]
[-5.6953279 -102.62655378 -166.89825218 -113.5671728 -74.77730011]
[-60.32758958 -156.12976174 -165.71728113 -141.27406351 -80.92644935]
[-112.50930892 -240.20593139 -157.93331919 -130.24323267 -46.59709758]
[-83.34281788 -215.41622939 -104.54387005 -27.18890459 0.]]
V(s):
[[0. -3.07397407 -6.5835953 -8.06843736 -8.36537886]
[-1.13906558 -7.89435029 -7.94753582 -7.57114485 -7.47773001]
[-5.02729913 -7.43475056 -8.28586406 -7.84855908 -7.35694994]
[-7.50062059 -8.00686438 -7.89666596 -8.68288218 -5.8246372]
[-8.33428179 -7.97837887 -6.53399188 -5.43778092 0.]]

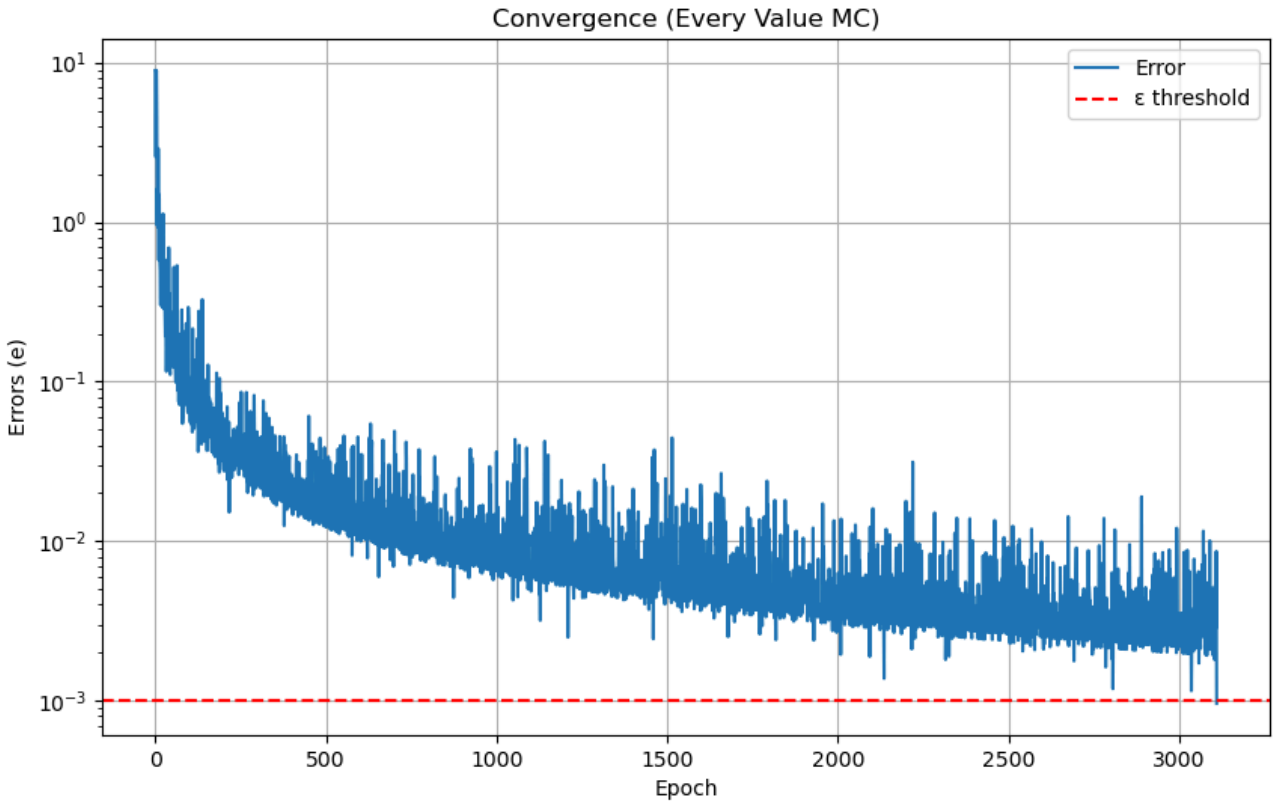
Returns for this episode:
k s r Y G(s)
1 (2, 3) -1 0.9 -9.202
2 (2, 2) -1 0.9 -9.114
3 (1, 2) -1 0.9 -9.015
4 (0, 2) -1 0.9 -8.906
5 (1, 2) -1 0.9 -8.784
6 (1, 1) -1 0.9 -8.649
7 (1, 2) -1 0.9 -8.499
8 (1, 1) -1 0.9 -8.332
9 (1, 2) -1 0.9 -8.147
10 (2, 2) -1 0.9 -7.941
11 (2, 1) -1 0.9 -7.712
12 (3, 1) -1 0.9 -7.458
13 (3, 0) -1 0.9 -7.176
14 (2, 0) -1 0.9 -6.862
15 (3, 0) -1 0.9 -6.513
16 (4, 0) -1 0.9 -6.126
17 (4, 1) -1 0.9 -5.695
18 (4, 2) -1 0.9 -5.217
19 (3, 2) -1 0.9 -4.686
20 (4, 2) -1 0.9 -4.095
21 (4, 1) -1 0.9 -3.439
22 (3, 1) -1 0.9 -2.710
23 (3, 2) -1 0.9 -1.900

24 (4, 2) -1 0.9 -1.000
25 (4, 3) 0 0.9 0.000

Epoch 3109
N(s):
[[0. 2357. 3536. 4078. 2920.]
[2377. 4220. 5037. 5354. 4228.]
[3666. 5001. 5238. 5133. 3752.]
[4371. 5549. 5126. 4228. 2488.]
[2950. 4157. 3570. 2330. 0.]]
S(s):
[[0. -11767.09219778 -25576.23871657 -33106.21886485
-24160.02253798]
[-11758.26439363 -28556.18432138 -39170.65125966 -43713.05842833
-34232.10947]
[-26561.06136851 -38780.95801219 -41973.12273601 -39965.9695462
-27200.18292395]
[-35383.83721509 -45422.79854216 -40104.70517414 -28276.37671874
-11921.78053281]
[-24372.83698616 -33696.24935218 -25789.4823312 -11470.15882113
0.]]
V(s):
[[0. -4.99240229 -7.23309918 -8.11824886 -8.27398032]
[-4.94668254 -6.76686832 -7.77658353 -8.16456078 -8.09652542]
[-7.24524314 -7.75464067 -8.0131964 -7.78608407 -7.2495157]
[-8.09513549 -8.18576294 -7.82378174 -6.68788475 -4.79171243]
[-8.26197864 -8.10590555 -7.22394463 -4.92281494 0.]]

Returns for this episode:
k s r Y G(s)
1 (2, 3) -1 0.9 -9.775
2 (1, 3) -1 0.9 -9.750
3 (2, 3) -1 0.9 -9.722
4 (2, 2) -1 0.9 -9.691
5 (3, 2) -1 0.9 -9.657
6 (2, 2) -1 0.9 -9.618
7 (2, 1) -1 0.9 -9.576
8 (2, 0) -1 0.9 -9.529
9 (2, 1) -1 0.9 -9.477
10 (2, 0) -1 0.9 -9.419
11 (3, 0) -1 0.9 -9.354

12 (2, 0) -1 0.9 -9.282
13 (2, 1) -1 0.9 -9.202
14 (3, 1) -1 0.9 -9.114
15 (3, 0) -1 0.9 -9.015
16 (3, 1) -1 0.9 -8.906
17 (4, 1) -1 0.9 -8.784
18 (3, 1) -1 0.9 -8.649
19 (2, 1) -1 0.9 -8.499
20 (2, 0) -1 0.9 -8.332
21 (1, 0) -1 0.9 -8.147
22 (1, 1) -1 0.9 -7.941
23 (2, 1) -1 0.9 -7.712
24 (2, 0) -1 0.9 -7.458
25 (2, 1) -1 0.9 -7.176
26 (2, 0) -1 0.9 -6.862
27 (2, 1) -1 0.9 -6.513
28 (2, 2) -1 0.9 -6.126
29 (2, 3) -1 0.9 -5.695
30 (2, 2) -1 0.9 -5.217
31 (3, 2) -1 0.9 -4.686
32 (4, 2) -1 0.9 -4.095
33 (4, 1) -1 0.9 -3.439
34 (3, 1) -1 0.9 -2.710
35 (2, 1) -1 0.9 -1.900
36 (2, 0) -1 0.9 -1.000
37 (1, 0) 0 0.9 0.000



---------- Part 3 Montecarlo On Policy ----------

Epoch 0
N(s):
[[0. 2. 2. 4. 7.]
[2. 4. 1. 3. 8.]
[7. 10. 5. 2. 3.]
[6. 8. 5. 0. 1.]
[3. 4. 2. 0. 0.]]
S(s):
[[0. -9.96618608 -19.65339743 -39.41909432 -69.19080018]
[-12.2085435 -21.19046258 -9.65663162 -28.80319796 -77.85126742]
[-62.05463445 -75.28720386 -45.96686683 -18.59131752 -28.22959177]
[-59.98747212 -68.92588791 -46.46907892 0. -9.47665237]
[-29.99201314 -38.45898265 -18.6291815 0. 0.]]
V(s):
[[0. -4.98309304 -9.82669872 -9.85477358 -9.88440003]
[-6.10427175 -5.29761565 -9.65663162 -9.60106599 -9.73140843]
[-8.86494778 -7.52872039 -9.19337337 -9.29565876 -9.40986392]
[-9.99791202 -8.61573599 -9.29381578 0. -9.47665237]
[-9.99733771 -9.61474566 -9.31459075 0. 0.]]

Returns for this episode:

k s r Y G(s)

1 (3, 0) -1 0.9 -9.999

2 (2, 0) -1 0.9 -9.999

3 (3, 0) -1 0.9 -9.999

4 (4, 0) -1 0.9 -9.999

5 (3, 0) -1 0.9 -9.999

6 (3, 1) -1 0.9 -9.998

7 (3, 2) -1 0.9 -9.998

8 (3, 1) -1 0.9 -9.998

9 (3, 0) -1 0.9 -9.998

10 (2, 0) -1 0.9 -9.998

11 (3, 0) -1 0.9 -9.997

12 (4, 0) -1 0.9 -9.997

13 (4, 1) -1 0.9 -9.997

14 (4, 0) -1 0.9 -9.996

15 (3, 0) -1 0.9 -9.996

16 (2, 0) -1 0.9 -9.995

17 (2, 1) -1 0.9 -9.995

18 (2, 0) -1 0.9 -9.994

19 (2, 1) -1 0.9 -9.994

20 (2, 2) -1 0.9 -9.993

21 (3, 2) -1 0.9 -9.992

22 (2, 2) -1 0.9 -9.991

23 (2, 1) -1 0.9 -9.990

24 (2, 0) -1 0.9 -9.989

25 (2, 1) -1 0.9 -9.988

26 (3, 1) -1 0.9 -9.987

27 (4, 1) -1 0.9 -9.985

28 (3, 1) -1 0.9 -9.984

29 (3, 2) -1 0.9 -9.982

30 (4, 2) -1 0.9 -9.980

31 (4, 1) -1 0.9 -9.978

32 (3, 1) -1 0.9 -9.975

33 (2, 1) -1 0.9 -9.973

34 (1, 1) -1 0.9 -9.970

35 (0, 1) -1 0.9 -9.966

36 (0, 2) -1 0.9 -9.962

37 (0, 3) -1 0.9 -9.958

38 (0, 4) -1 0.9 -9.954

39 (1, 4) -1 0.9 -9.948
40 (0, 4) -1 0.9 -9.943
41 (1, 4) -1 0.9 -9.936
42 (0, 4) -1 0.9 -9.929
43 (0, 3) -1 0.9 -9.921
44 (0, 4) -1 0.9 -9.913
45 (1, 4) -1 0.9 -9.903
46 (1, 3) -1 0.9 -9.892
47 (1, 4) -1 0.9 -9.880
48 (0, 4) -1 0.9 -9.867
49 (1, 4) -1 0.9 -9.852
50 (0, 4) -1 0.9 -9.836
51 (0, 3) -1 0.9 -9.818
52 (1, 3) -1 0.9 -9.797
53 (1, 4) -1 0.9 -9.775
54 (0, 4) -1 0.9 -9.750
55 (0, 3) -1 0.9 -9.722
56 (0, 2) -1 0.9 -9.691
57 (1, 2) -1 0.9 -9.657
58 (2, 2) -1 0.9 -9.618
59 (2, 3) -1 0.9 -9.576
60 (2, 4) -1 0.9 -9.529
61 (3, 4) -1 0.9 -9.477
62 (2, 4) -1 0.9 -9.419
63 (1, 4) -1 0.9 -9.354
64 (2, 4) -1 0.9 -9.282
65 (1, 4) -1 0.9 -9.202
66 (1, 3) -1 0.9 -9.114
67 (2, 3) -1 0.9 -9.015
68 (2, 2) -1 0.9 -8.906
69 (3, 2) -1 0.9 -8.784
70 (4, 2) -1 0.9 -8.649
71 (4, 1) -1 0.9 -8.499
72 (3, 1) -1 0.9 -8.332
73 (2, 1) -1 0.9 -8.147
74 (3, 1) -1 0.9 -7.941
75 (3, 2) -1 0.9 -7.712
76 (2, 2) -1 0.9 -7.458
77 (2, 1) -1 0.9 -7.176
78 (2, 0) -1 0.9 -6.862

79 (1, 0) -1 0.9 -6.513
80 (1, 1) -1 0.9 -6.126
81 (1, 0) -1 0.9 -5.695
82 (2, 0) -1 0.9 -5.217
83 (2, 1) -1 0.9 -4.686
84 (1, 1) -1 0.9 -4.095
85 (2, 1) -1 0.9 -3.439
86 (3, 1) -1 0.9 -2.710
87 (2, 1) -1 0.9 -1.900
88 (1, 1) -1 0.9 -1.000
89 (0, 1) 0 0.9 0.000

Epoch 1
N(s):
[[0. 2. 2. 4. 7.]
[3. 6. 1. 3. 8.]
[7. 12. 6. 3. 3.]
[6. 8. 6. 1. 1.]
[3. 4. 2. 0. 0.]]
S(s):
[[0. -9.96618608 -19.65339743 -39.41909432 -69.19080018]
[-12.2085435 -24.90046258 -9.65663162 -28.80319796 -77.85126742]
[-62.05463445 -80.62620386 -50.06196683 -23.27690752 -28.22959177]
[-59.98747212 -68.92588791 -52.16440682 -5.217031 -9.47665237]
[-29.99201314 -38.45898265 -18.6291815 0. 0.]]
V(s):
[[0. -4.98309304 -9.82669872 -9.85477358 -9.88440003]
[-4.0695145 -4.1500771 -9.65663162 -9.60106599 -9.73140843]
[-8.86494778 -6.71885032 -8.34366114 -7.75896917 -9.40986392]
[-9.99791202 -8.61573599 -8.6940678 -5.217031 -9.47665237]
[-9.99733771 -9.61474566 -9.31459075 0. 0.]]

Returns for this episode:
k s r Y G(s)
1 (3, 2) -1 0.9 -5.695
2 (3, 3) -1 0.9 -5.217
3 (2, 3) -1 0.9 -4.686
4 (2, 2) -1 0.9 -4.095
5 (2, 1) -1 0.9 -3.439
6 (1, 1) -1 0.9 -2.710
7 (2, 1) -1 0.9 -1.900

8 (1, 1) -1 0.9 -1.000
9 (1, 0) 0 0.9 0.000

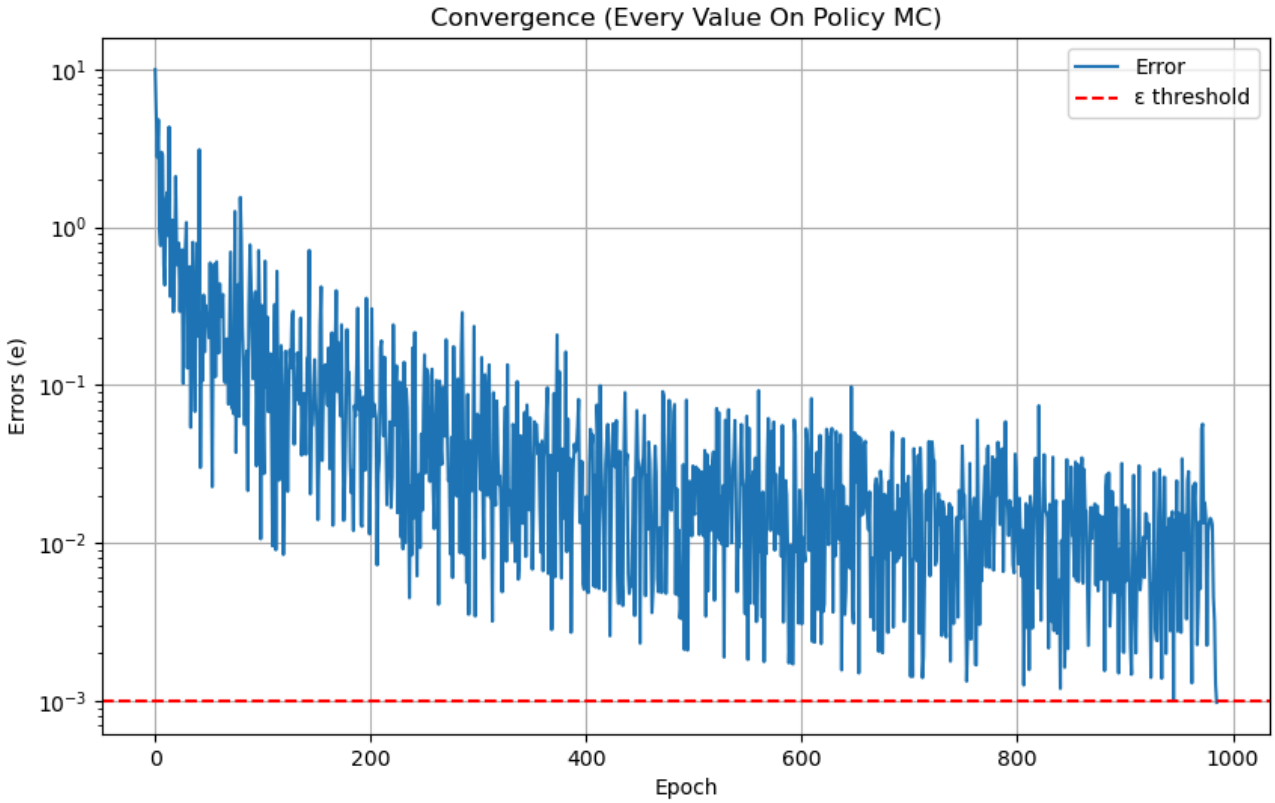
Epoch 10
N(s):
[[0. 2. 2. 5. 7.]
[6. 7. 1. 4. 8.]
[9. 14. 6. 5. 3.]
[6. 9. 8. 9. 4.]
[6. 9. 5. 8. 0.]]
S(s):
[[0. -9.96618608 -19.65339743 -42.85809432 -69.19080018]
[-12.2085435 -25.90046258 -9.65663162 -31.51319796 -77.85126742]
[-64.05463445 -84.42620386 -50.06196683 -33.32388733 -28.22959177]
[-59.98747212 -71.63588791 -69.44769392 -56.80240411 -17.18897312]
[-42.01414414 -53.82257265 -24.7242815 -24.06983813 0.]]
V(s):
[[0. -4.98309304 -9.82669872 -8.57161886 -9.88440003]
[-2.03475725 -3.70006608 -9.65663162 -7.87829949 -9.73140843]
[-7.11718161 -6.03044313 -8.34366114 -6.66477747 -9.40986392]
[-9.99791202 -7.9595431 -8.68096174 -6.31137823 -4.29724328]
[-7.00235736 -5.98028585 -4.9448563 -3.00872977 0.]]

Returns for this episode:
k s r Y G(s)
1 (3, 4) 0 0.9 0.000

Epoch 985
N(s):
[[0. 243. 184. 125. 65.]
[332. 259. 161. 84. 72.]
[131. 79. 55. 88. 113.]
[67. 87. 102. 183. 141.]
[78. 204. 347. 481. 0.]]
S(s):
[[0. -115.04438728 -335.90300779 -389.98384155 -292.91438947]
[-140.0718407 -434.93205299 -441.48509625 -310.73389432 -264.50205249]
[-270.10865845 -260.8279564 -222.3023033 -266.43317388 -239.19978198]
[-233.99104466 -340.59295281 -328.54331786 -345.92668522 -100.48536872]
[-295.39954051 -598.16550201 -612.71542289 -224.95391973 0.]]
V(s):

[[0. -0.47343369 -1.82555982 -3.11987073 -4.50637522]
[-0.42190313 -1.67927434 -2.74214345 -3.69921303 -3.67363962]
[-2.06189816 -3.3016197 -4.04186006 -3.0276497 -2.11681223]
[-3.49240365 -3.91486153 -3.22101292 -1.89030976 -0.71266219]
[-3.7871736 -2.93218383 -1.7657505 -0.46767967 0.]]

Returns for this episode:
k s r Y G(s)
1 (4, 3) 0 0.9 0.000



---------- Part 4 Standard Q Learning----------
Rewards Matrix (R):
[[100 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 100]]

Q-matrix at epoch 0:
[[[0. 0. 0. 0.]
[0. -1. 0. -1.]
[0. 0. -1. -1.]
[0. -1. -1. -1.]
[0. 0. 0. -1.]]

[[0. -1.9 0. 0.]
[-1. -1. -1. -1.]
[0. -1. 0. -1.]
[-1. -1. -1.9 0.]
[-1. 0. 0. -1.]]

[[-1. 0. -1. 0.]
[-1. -1.9 -1.9 -1.]
[-1. -1.9 -1. -1.9]
[-1. -1. -1. -1.]
[-1. 0. -1. -1.9]]

[[0. -1.9 -1. 0.]
[-1.9 -1. -1. -1.]
[-1. -1. 0. 0.]
[-1. -1. -1. -1.]
[-1. 0. 0. -1.]]

[[-1. -1. 0. 0.]
[-1.9 0. 0. -1.]
[0. 0. 0. -1.]
[0. 0. 0. -1.]
[0. 0. 0. 0.]]]

Q-matrix at epoch 1:
[[[0. 0. 0. 0.]
[0. -1. 0. -1.]
[0. -1.9 -1. -1.]
[0. -1.9 -1. -1.]
[0. 0. -1.9 -1.]]

[[0. -1.9 0. 0.]
[-1. -1. -1. -1.]
[-1.9 -1. 0. -1.]
[-1.9 -1. -1.9 -1.]
[-1.9 0. -1.9 -1.]]

[[-1. 0. -1. 0.]
[-1. -1.9 -1.9 -1.]
[-1. -1.9 -1. -1.9]
[-1. -1.9 -1. -1.]
[-1.9 0. -1. -1.9]]

[[0. -1.9 -1. 0.]
[-1.9 -1. -1. -1.]
[-1. -1. 0. 0.]
[-1. -1. -1. -1.]
[-1. 0. 0. -1.]]

[[-1. -1. 0. 0.]
[-1.9 0. 0. -1.]
[0. 0. 0. -1.]
[0. 0. 0. -1.]
[0. 0. 0. 0.]]]

Q-matrix at epoch 10:
[[[0. 0. 0. 0.]
[0. -2.71 -1.9 -1.]
[0. -2.71 -1.9 -1.9]
[0. -1.9 -1.9 -2.71]
[0. 0. -2.71 -1.9]]

[[-1. -1.9 0. 0.]
[-1.9 -1.9 -2.71 -1.]
[-2.71 -1. -1.9 -1.9]
[-1.9 -2.71 -2.71 -1.9]
[-1.9 0. -2.71 -2.71]]

[[-1. -2.71 -1.9 0.]
[-1.9 -1.9 -1.9 -1.9]
[-1.9 -2.71 -1. -2.71]
[-2.71 -1.9 -1.9 -1.9]
[-2.71 0. -1.9 -1.9]]

[[-1.9 -1.9 -2.71 0.]
[-2.71 -1. -2.71 -1.9]
[-1.9 -1. -1. -1.9]
[-1.9 -1. -1. -1.]
[-2.71 0. -1. -1.]]

[[-2.71 -2.71 0. 0.]
[-1.9 -1.9 0. -2.71]
[-1.9 -1. 0. -1.]
[0. -1. 0. -1.]
[0. 0. 0. 0.]]]

Converged at epoch 52

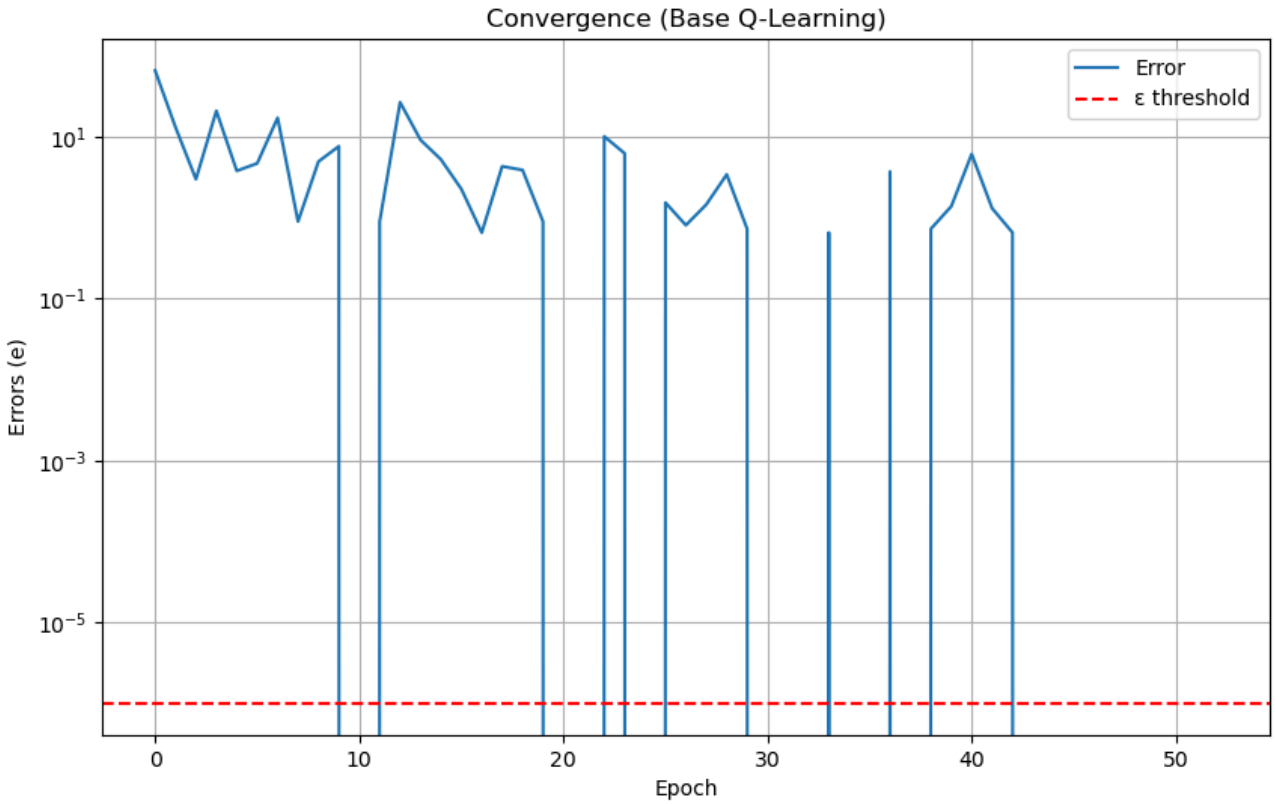
Final Q-matrix:
[[[0. 0. 0. 0.]
[0. -2.71 -2.71 -1.]
[0. -3.439 -3.439 -1.9]
[0. -4.0951 -4.0951 -2.71]
[0. 0. -3.439 -3.439]]

[[-1. -2.71 -2.71 0.]
[-1.9 -3.439 -3.439 -1.9]
[-2.71 -4.0951 -4.0951 -2.71]
[-3.439 -3.439 -3.439 -3.439]
[-4.0951 0. -2.71 -4.0951]]

[[-1.9 -3.439 -3.439 0.]
[-2.71 -4.0951 -4.0951 -2.71]
[-3.439 -3.439 -3.439 -3.439]
[-4.0951 -2.71 -2.71 -4.0951]
[-3.439 0. -1.9 -3.439]]

[[-2.71 -4.0951 -4.0951 0.]
[-3.439 -3.439 -3.439 -3.439]
[-4.0951 -2.71 -2.71 -4.0951]
[-3.439 -1.9 -1.9 -3.439]
[-2.71 0. -1. -2.71]]

[[-3.439 -3.439 0. 0.]
[-4.0951 -2.71 0. -4.0951]
[-3.439 -1.9 0. -3.439]
[-2.71 -1. 0. -2.71]
[0. 0. 0. 0.]]]
[['.' '<' '<' '<' 'V']
['^' '^' '^' '^' 'V']
['^' '^' '^' '>' 'V']
['^' '^' '>' '>' 'V']
['^' '>' '>' '>' '.']]



---------- Part 5 SARSA ----------

Rewards Matrix (R):

[[100 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 100]]

Q-matrix at epoch 0:

[[[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]]

[[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]]

[[0. -1. -1. 0.]
[0. 0. 0. -1.]

[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]]

[[-1. -1. 0. 0.]
[0. 0. -1. 0.]
[0. -1. 0. 0.]
[0. -1. 0. 0.]
[0. 0. -1. 0.]]

[[-1. -1. 0. 0.]
[0. -1. 0. -1.]
[-1. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]]]

Q-matrix at epoch 1:
[[[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]]

[[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. -1. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]]

[[0. -1. -1. 0.]
[0. -1. -1.9 -1.]
[-1. -1. -1. -1.]
[0. 0. 0. -1.]
[0. 0. 0. 0.]]

[[-1. -1.9 -1.9 0.]
[-1. -1. -1. -1.9]
[-1. -1. -1. -1.]
[0. -1. 0. 0.]
[0. 0. -1. 0.]]

[[-1. -1.9 0. 0.]
[-1. -1. 0. -1.]

[-1. -1. 0. -1.9]
[0. -1. 0. 0.]
[0. 0. 0. 0.]]]

Q-matrix at epoch 10:
[[[0. 0. 0. 0.]
[0. -1.9 -1.9 -1.]
[0. -1. -1.9 -1.9]
[0. -1. -1.9 -1.9]
[0. 0. -1.9 -1.9]]

[[-1. -1.9 -2.71 0.]
[-1.9 -1.9 -1.9 -1.9]
[-1.9 -1.9 -2.71 -1.9]
[-1. -1.9 -1.9 -1.9]
[-1.9 0. -1.9 -1.]]

[[-1.9 -1.9 -1.9 0.]
[-2.71 -2.71 -1.9 -2.71]
[-2.71 -1.9 -2.71 -2.71]
[-1.9 -1.9 -1.9 -1.9]
[-1. 0. -1. -1.9]]

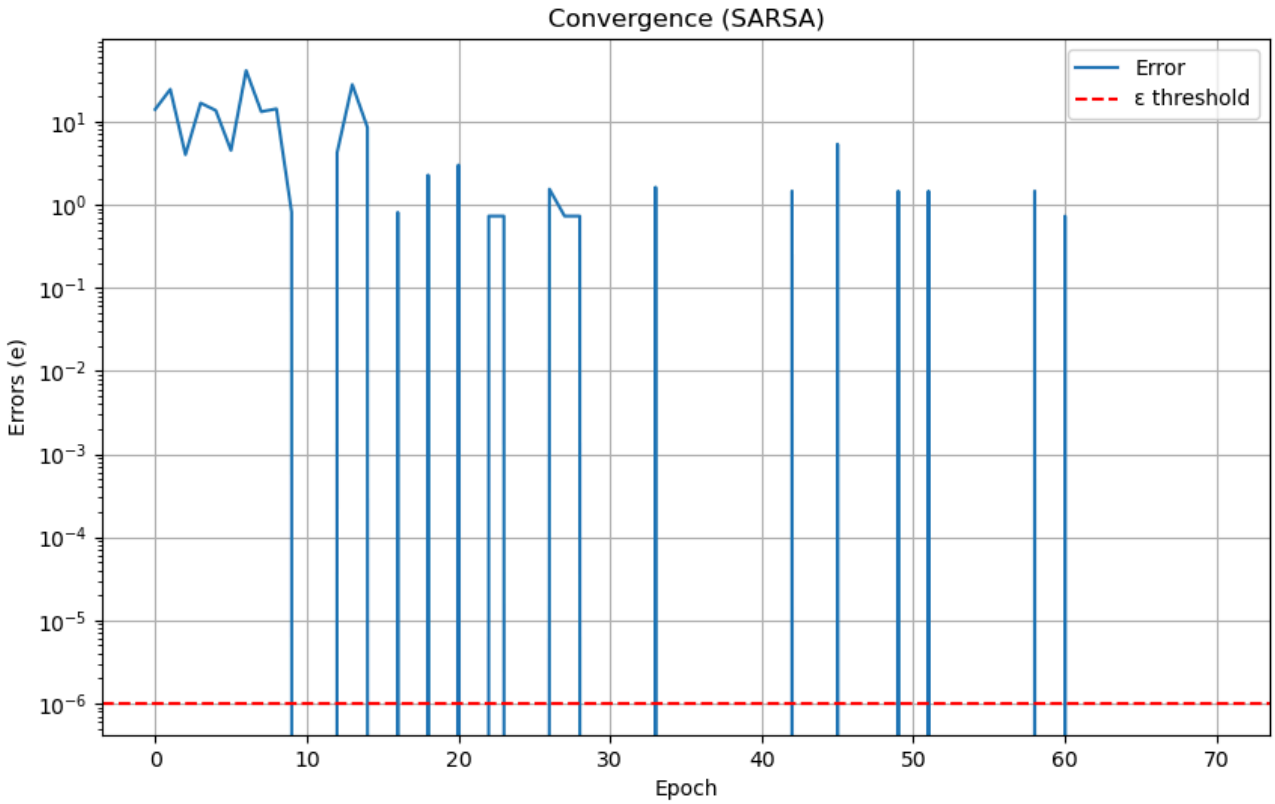
[[-2.71 -2.71 -1.9 0.]
[-2.71 -2.71 -1.9 -2.71]
[-1.9 -1.9 -1.9 -2.71]
[-1.9 -1.9 -1. -1.9]
[-1.9 0. -1. -1.]]

[[-2.71 -1.9 0. 0.]
[-1.9 -2.71 0. -2.71]
[-1.9 -1.9 0. -1.9]
[-1. -1. 0. -1.9]
[0. 0. 0. 0.]]]

Converged at epoch 70

Final Q-matrix:
[[[0. 0. 0. 0.]
[0. -1.9 -1.9 -1.]
[0. -3.439 -3.439 -1.9]
[0. -3.439 -2.71 -2.71]
[0. 0. -3.439 -3.439]]

```
[[-1. -1.9 -2.71 0. ]
[-1.9 -2.71 -3.439 -1.9 ]
[-2.71 -2.71 -3.439 -2.71 ]
[-3.439 -3.439 -3.439 -3.439]
[-3.439 0. -2.71 -3.439]]

[[-1.9 -2.71 -3.439 0. ]
[-2.71 -2.71 -3.439 -2.71 ]
[-3.439 -3.439 -3.439 -3.439]
[-3.439 -2.71 -2.71 -3.439]
[-2.71 0. -1.9 -2.71 ]]

[[-2.71 -3.439 -3.439 0. ]
[-3.439 -3.439 -3.439 -3.439]
[-2.71 -2.71 -2.71 -2.71 ]
[-2.71 -1.9 -1.9 -2.71 ]
[-1.9 0. -1. -1.9 ]]

[[-3.439 -3.439 0. 0. ]
[-3.439 -2.71 0. -3.439]
[-2.71 -1.9 0. -3.439]
[-2.71 -1. 0. -1.9 ]
[ 0. 0. 0. 0. ]]]
[['.' '<' '<' '>' 'V']
['^' '^' '^' '^' 'V']
['^' '^' '^' '>' 'V']
['^' '^' '^' '>' 'V']
['^' '>' '>' '>' '.']]
```



---------- Part 6 Epsilon Greedy ----------

Rewards Matrix (R):

[[100 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 -1]
[-1 -1 -1 -1 100]]

Q-matrix at epoch 0:

[[[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. -1. 0. 0.]
[0. -1. -1. 0.]
[0. 0. -1.9 0.]]

[[-1. 0. 0. 0.]
[0. 0. 0. -1.]
[-1. 0. 0. -1.]
[-1. 0. 0. -1.]
[-1. 0. -1. -1.]]

[[0. 0. 0. 0.]
[0. 0. 0. 0.]

[0. 0. 0. 0.]
[-1. -1. 0. 0.]
[-1. 0. -1. -1.]]

[[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. -1. 0. 0.]
[-1.9 0. 0. -1.]]

[[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]]]

Q-matrix at epoch 1:
[[[0. 0. 0. 0.]
[0. -1. 0. 0.]
[0. -1. -1. 0.]
[0. -1. -1. 0.]
[0. 0. -1.9 0.]]

[[-1. 0. 0. 0.]
[-1. -1. -1. -1.]
[-1. 0. 0. -1.]
[-1. 0. 0. -1.]
[-1. 0. -1. -1.]]

[[0. 0. 0. 0.]
[-1. 0. -1.9 0.]
[0. 0. 0. 0.]
[-1. -1. 0. 0.]
[-1. 0. -1. -1.]]

[[0. -1. -1. 0.]
[-1. -1. -1. -1.]
[0. 0. 0. -1.]
[0. -1. -1. 0.]
[-1.9 0. 0. -1.]]

[[0. -1. 0. 0.]
[-1. -1.9 0. 0.]

[-1. -1. 0. -1.]
[-1. -1. 0. -1.]
[0. 0. 0. 0.]]]

Q-matrix at epoch 10:
[[[0. 0. 0. 0.]
[0. -1. -1.9 -1.]
[0. -1.9 -2.71 -1.9]
[0. -2.71 -1.9 -1.9]
[0. 0. -2.71 -1.9]]

[[-1. -1.9 -1. 0.]
[-1. -2.71 -1.9 -1.9]
[-1.9 -1.9 -2.71 -1.9]
[-2.71 -1.9 -2.71 -2.71]
[-1.9 0. -1.9 -2.71]]

[[-1.9 -1.9 -1.9 0.]
[-1.9 -1.9 -2.71 -1.9]
[-2.71 -2.71 -2.71 -2.71]
[-1.9 -1.9 -1.9 -1.9]
[-2.71 0. -1. -1.]]

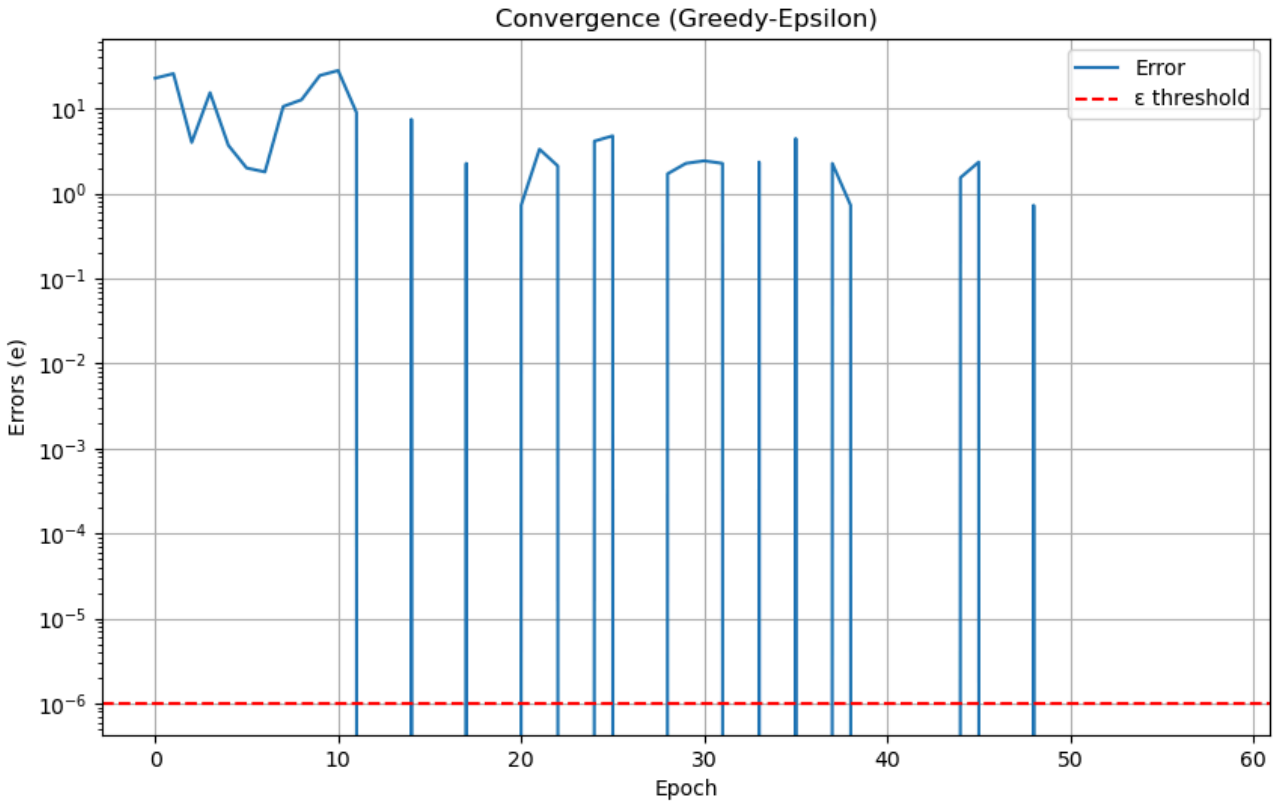
[[-2.71 -1.9 -1. 0.]
[-2.71 -2.71 -1.9 -1.9]
[-2.71 -1.9 -1.9 -2.71]
[-1.9 -1.9 -1.9 -2.71]
[-1.9 0. -1. -1.]]

[[-1.9 -1.9 0. 0.]
[-1.9 -1.9 0. -1.9]
[-1.9 -1.9 0. -1.9]
[-1. -1. 0. -2.71]
[0. 0. 0. 0.]]]

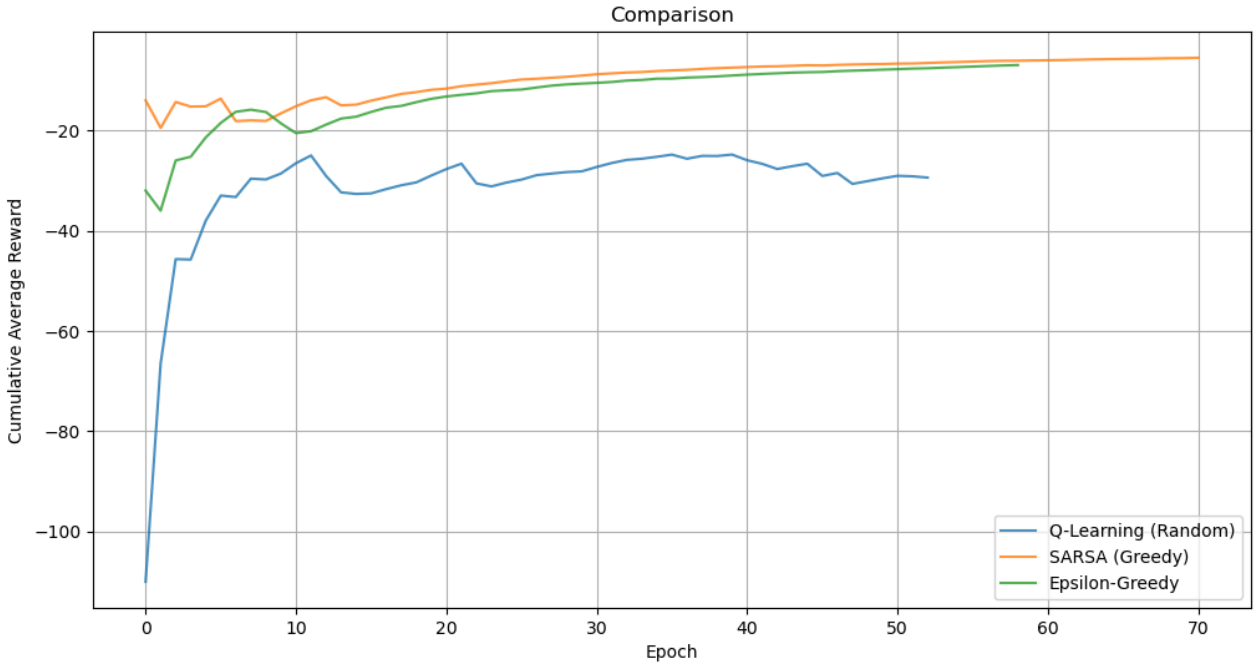
Converged at epoch 58

Final Q-matrix:
[[[0. 0. 0. 0.]
[0. -2.71 -2.71 -1.]
[0. -2.71 -2.71 -1.9]
[0. -3.439 -3.439 -2.71]
[0. 0. -3.439 -3.439]]

```
[[-1. -1.9 -2.71 0. ]
[-1.9 -2.71 -1.9 -1.9 ]
[-2.71 -3.439 -2.71 -2.71 ]
[-3.439 -3.439 -3.439 -3.439 ]
[-4.0951 0. -2.71 -3.439 ]]

[[-1.9 -2.71 -2.71 0. ]
[-2.71 -3.439 -3.439 -2.71 ]
[-2.71 -3.439 -2.71 -3.439 ]
[-3.439 -2.71 -2.71 -3.439 ]
[-2.71 0. -1.9 -2.71 ]]

[[-2.71 -2.71 -2.71 0. ]
[-2.71 -3.439 -3.439 -3.439 ]
[-3.439 -2.71 -2.71 -2.71 ]
[-2.71 -1.9 -1.9 -2.71 ]
[-1.9 0. -1. -2.71 ]]

[[-3.439 -2.71 0. 0. ]
[-3.439 -2.71 0. -2.71 ]
[-2.71 -1.9 0. -3.439 ]
[-2.71 -1. 0. -2.71 ]
[ 0. 0. 0. 0. ]]]
[['.' '<' '<' '<' 'V']
['^' '^' '^' '^' 'V']
['^' '^' '^' '>' 'V']
['^' '^' '>' '>' 'V']
['>' '>' '>' '>' '.']]
```



---------- Part 7 Comparing Rewards ----------



Answering Questions

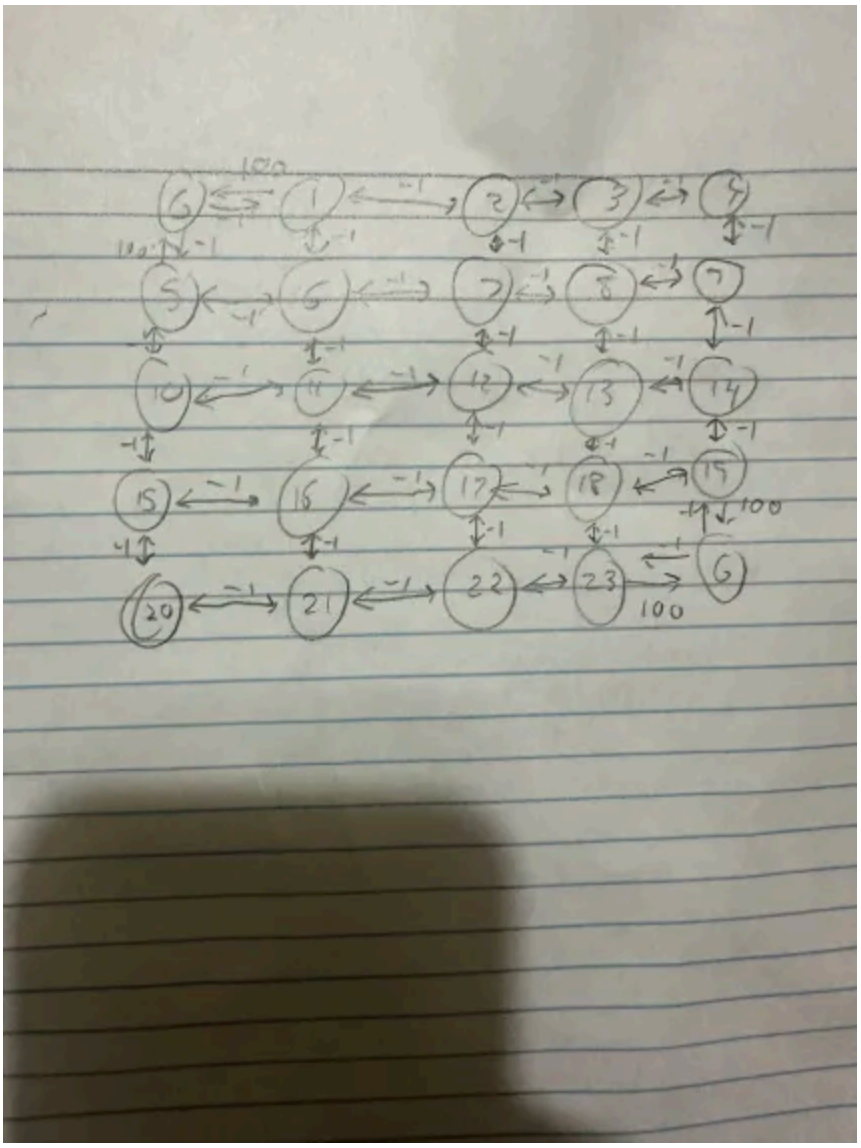
The previous stuff was all computer output, from here on it is me typing the answers to the questions

QUESTION 1: For all the MonteCarlo convergences, what I ended up doing was subtracting the new V(s) (values) subtracting the old ones from it, taking the absolute value of the subtraction and then summing it across all squares so: Sum(|V(s) - oldV(s)|). I then took this number and compared it against a threshold (arbitrary), in this case it was .001 but it could be anything. I chose this because it was a good way to know that nothing much was changing from the last one to the current one and it was really easy to implement.

QUESTION 2: In this case, the First value converged about 20 epochs or ~2% earlier which is not a lot at all. In general though, the First value seemed to converge a little faster than the every value (though not by much). I suspect this is because the first value only takes the first time it reached a square while every value takes every time; this combined with the random nature of the montecarlo likely lead to some more outliers and larger changes between previous and old causing first value to converge first

QUESTION 3: It is more like SARSA because when we get to a square, we go random until we get to a known policy. This is very similar to SARSA where if there is a known policy we take it. The drawback of this was that it ended up getting to an infinite loop sometimes so I had to make it more like greedy-epsilon where there is a chance that the known policy is taken but there is also a chance that a random one is picked

QUESTION 4:



QUESTION 5: What I did for all the QLearning convergence is having an arbitrary threshold with the same subtraction sum equation (again) but instead of checking if the threshold was met (this proved wildly inconsistent), I would check if the sum of the last 10 iterations was less than the threshold. This way outliers stop the program early. It was also efficient and really easy to code/adjust

QUESTION 6: Up, left, left. Optimal policy was found throughout the board

QUESTION 7: Same as q5

QUESTION 8: Same as q6, though this specific square was fine, other squares did not have the optimal policy probably because it discoverer a way that worked and didn't change it

QUESTION 9: No, SARSA took longer to converge than random Q learning. This is likely an outlier because the other few times I ran it this was not the case. More often than not the regular Qlearning was able to converge the slowest because it relied on randomness. Sarsa was generally fastest because it used the most optimal path nearly every time. This was the only case where regular Qlearning was the fastest and that might just be to random luck and this test case being pretty simple

QUestion 10: Same as q5

Question 11: Same as q6

Question 12: No, QLearning converged first with SARSA second and GE 3rd. This was a bit unusual as generally, SARSA finishes first with GE second and QL 3rd. SARSA usually finishes first because it is the most optimal and least exploratory while GE is a mix of exploratory and optimal which is why it usually finishes second. QL is the most exploratory which is why it usually finishes last but in this case it finished first. This might be to the inherent random nature finding a good seed that works optimally. ANother reason could be because SARSA and GE are both random and exploratory at first and usually take some time to become more optimal. This task may just not have been hard/long enough for the optimal part to really make a difference.

Question 13: The rewards are different based on the way they explore. Since QL is more random it has a lower initial reward and takes longer to converge. Since SARSA is fully greedy it has the highest initial reward and converges the fastest but sometimes has a lower max (though not in this case). Since GE is a mix of both it is like a happy medium starting between both of them

Question 14: there are many reasons the plots are different including epochs done, the task being very simple compared to the one in class, and the task being completely different.