Assignment2

Sen Wang
03692569

Exercise 1

Priors:

k	1	2	3	4
Prior	0.201	0.262	0.240	0.297

Means:

k	1	2	3	4
mean	-0.0147, -0.0796	0.0262, 0.0617	-0.0432, 0.0446	-0.0194, -0.0166

Covariance Matrix

k	sigma
1	0.00039, 0.00022 0.00022, 0.00013
2	0.0011, -0.0004 -0.0004, 0.0002
3	0.00017, 0.00026 0.00026, 0.00040
4	0.00074, -0.00059 -0.00059, 0.00061

Exercise 2:

Classification Results

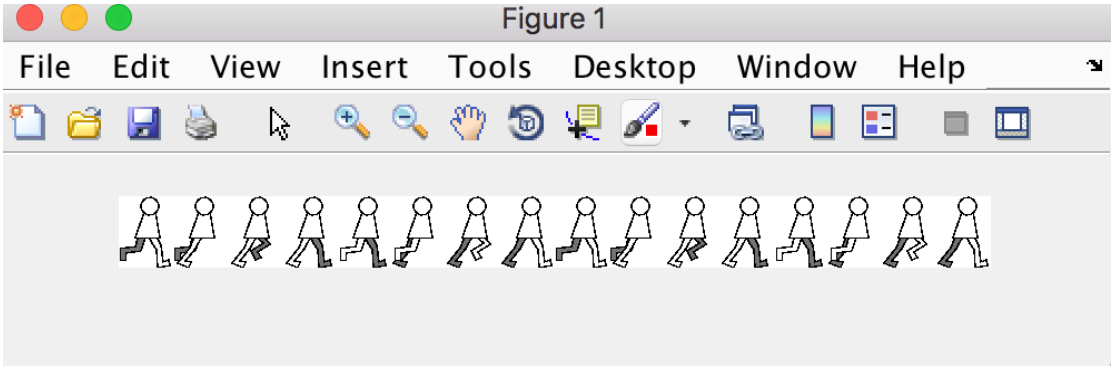
sequences	1	2	3	4	5	6	7	8	9	10
results	2	2	2	2	2	2	2	2	2	2

Exercise 3:

a)

Policy Iteration:

Initial State = 8



States: 8, 5, 9, 13, 14, 2, 3, 4, 8, 5, 9, 13, 14, 2, 3, 4

Policy Iteration:

1)

Reward Matrix

	A1	A2	A3	A4
S1	0	-5	0	-5
S2	0	0	-5	-5
S3	0	0	-5	-5
S4	0	-5	0	-5
S5	-5	-5	0	0
S6	0	0	0	0
S7	0	0	0	0
S8	-5	5	0	0
S9	-5	-5	0	0
S10	0	0	0	0
S11	0	0	0	0
S12	-5	5	0	0
S13	0	-5	0	-5
S14	0	0	-5	5
S15	0	0	-5	5
S16	0	5	0	5

2)

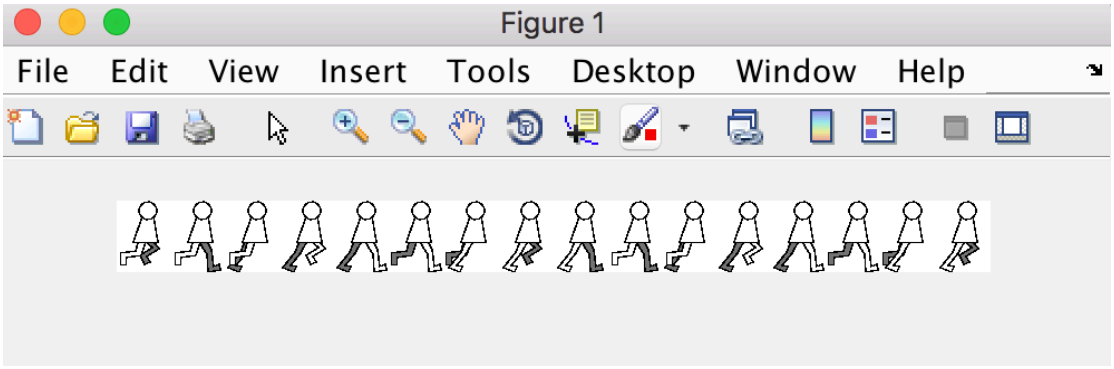
I have used γ=0.9, when γ increases, it takes more iterations to converge, when γ decreases, it takes less iterations to converge.

3)

Approximately it takes 7 iteration to converge.

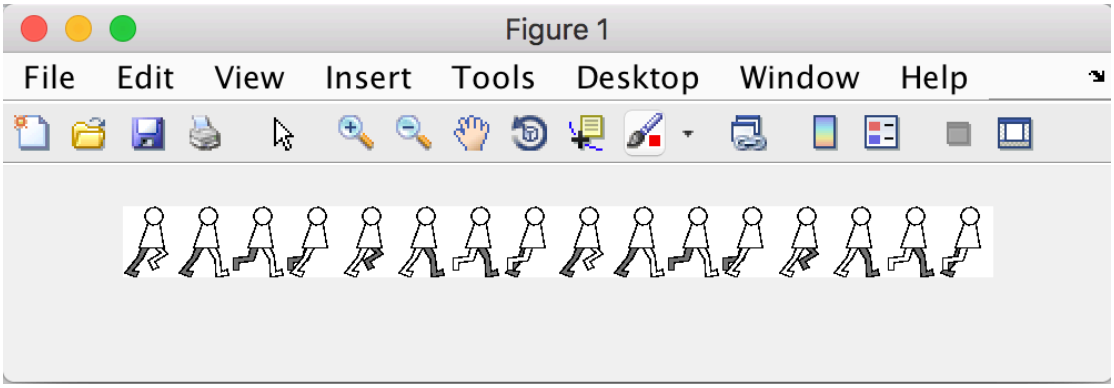
4)

Initial state = 10



states: 10, 14, 2, 3, 4, 8, 5, 9, 13, 14, 2, 3, 4, 8, 5, 9

Initial state = 3

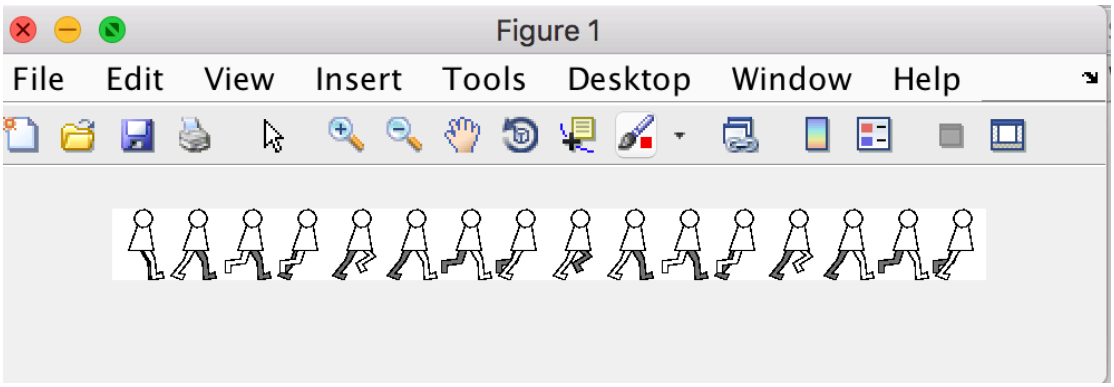


states: 3, 4, 8, 5, 9, 13, 14, 2, 3, 4, 8, 5, 9, 13, 14, 2

b)

Q learning:

Initial State = 16

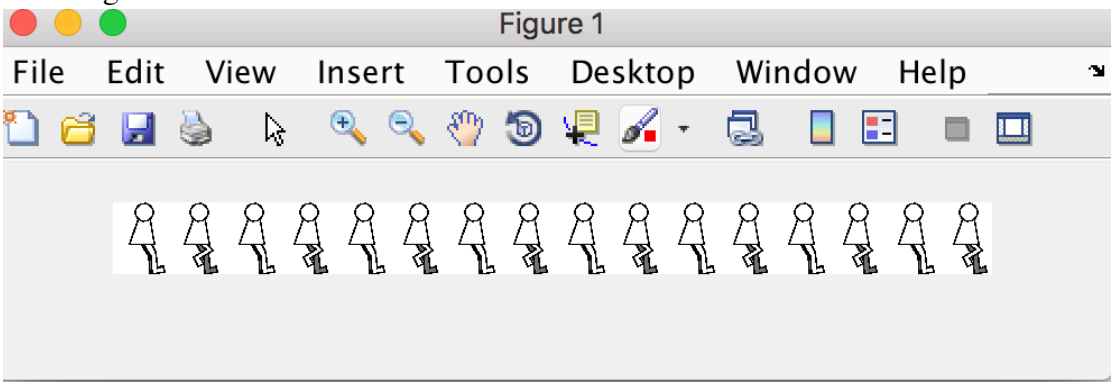


states: 16, 13, 14, 2, 3, 4, 8, 5, 9, 13, 14, 2, 3, 4, 8, 5

1)

$\epsilon = 0.01$ $\alpha = 0.6$

2)
If a pure greedy policy is used, then it will not converged at all, which can be shown in the figure below.

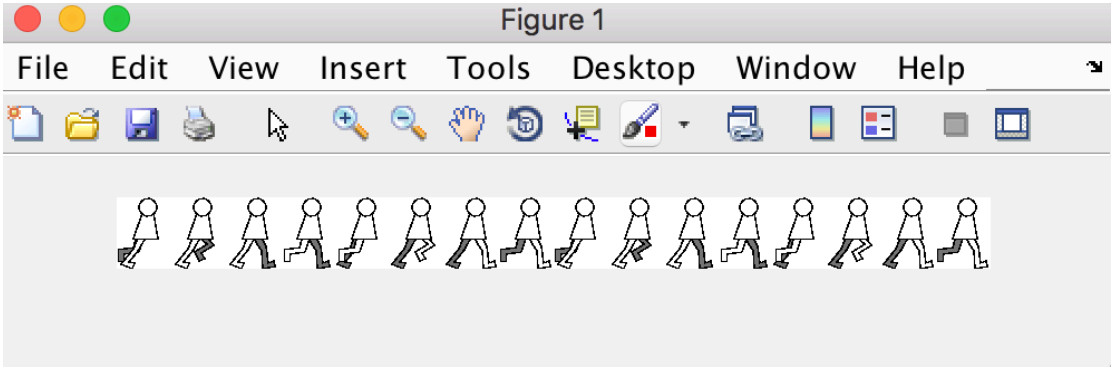


the value of ϵ will not affect the final result, but will affect the steps of convergence.

3)
Approxmately it takes 4500 – 5000 steps to get the optimal policy.

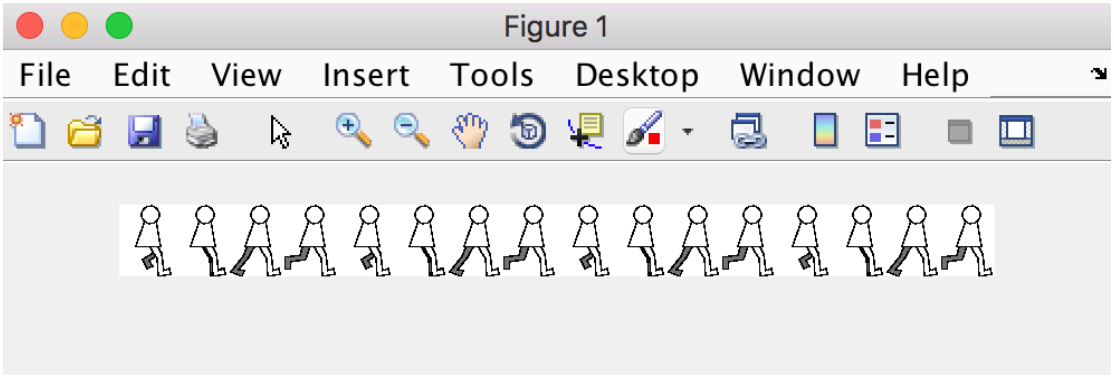
4)

Initial State: 5



states: 5, 9, 13, 14, 2, 3, 4, 8, 5, 9, 13, 14, 2, 3, 4, 8

Initial State: 12



states: 12, 16, 4, 8, 12, 16, 4, 8, 12, 16, 4, 8, 12, 16, 4, 8