Intelligent 12/13 Assignment 3 - Reinforcement Learning
Suttipong Mungkala mungks@usi.ch

Note that my codes are written in PHP and it has been tested running on a web browser. In order to run my codes you need PHP and web server(e.g. Apache) installed and configured. The instruction to setting up the environment to test my codes will be sent to you later by email.

Question 1.

A. (40 points) Implement Policy Evaluation. Starting with V (s) = 0, 8s, and assuming a random policy ((s, a) = 1/4, 8s, a), what are the final values, V (s), after the evaluation has converged?

The final values V(s) are shown as below:

"-1" is the walls.
"0" is the Goal state, which has not been updated.

0.018724	0.031374	0.070633	0.141305	0.238169	0.385376	0.534735
0.01443	0.018731	-1	0.177931	0.293691	0.554515	0.921764
0.012271	-1	-1	0.177889	0.334688	0.863686	2.085721
0.015588	0.02537	0.048108	0.100121	0.152241	-1	5.398703
0.016068	0.023707	0.040227	0.066753	0.089585	-1	0

B. (10 points) How and why do the values change if the discount factor, is changed to 0.7 (again starting with V (s) = 0, 8s)?

How - the evaluation has converged using a less episode (faster).
Why - then the agent cares more about selecting actions that maximize immediate reward, which is called *shortsighted.*

Question 2.

A. (30 points) Implement Q-learning with learning rate ↵ = 0.4. Initialize Q(s, a) = 0, 8s, a. Starting each episode in state S, run Q-learning until it converges, using an e-greedy policy. Each episode ends after 100 actions or once the goal, G, has been reached, whichever happens first.

The maze is represent as:

"-1" is the walls.
"30" is the Goal state
each number represent the state.

1, 6, 10, 13, 18, 23, 26
2, 7, -1, 14, 19, 24, 27
3, -1, -1, 15, 20, 25, 28
4, 8, 11, 16, 21, -1, 29
5, 9, 12, 17, 22, -1, 30

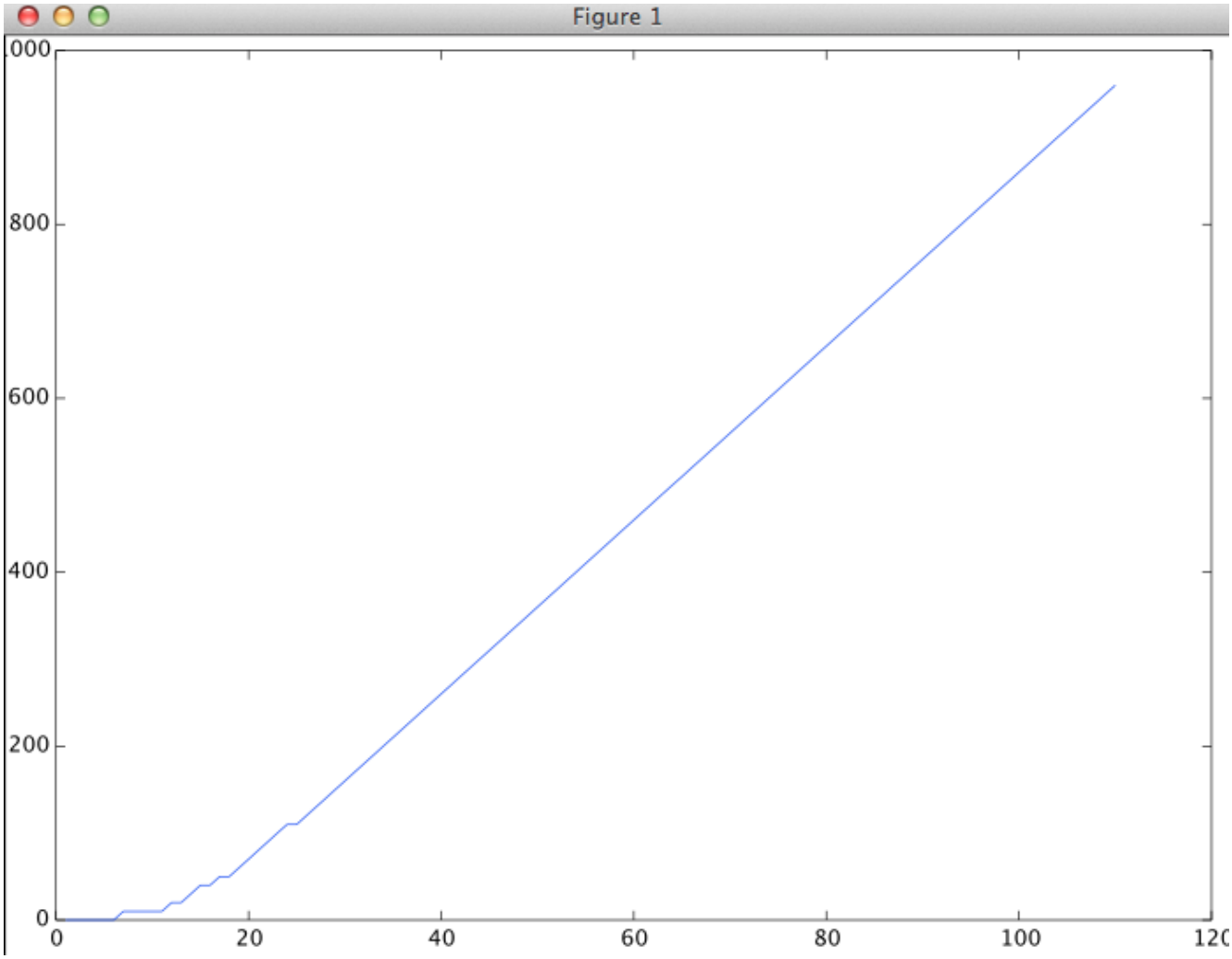
The result after convergence (1934 runs):

State	N	E	W	S
1	0	3.87	0	3.14
2	3.49	3.49	0	2.82
3	3.14	0	0	2.49
4	2.82	0	0	0
5	0.61	0	0	0
6	0	4.3	0	3.49
7	3.87	0	3.14	0
8	0	0	0	0
9	0	0	0	0
10	0	4.78	0	0
11	0	2.43	0	0
12	0	1.02	0	0
13	0	5.31	0	5.31
14	4.78	5.9	0	5.9
15	5.31	6.56	0	5.31
16	5.9	0	0.3	3.2
17	0	4.91	0	0
18	0	5.9	0	5.9
19	5.31	6.56	4.13	6.56
20	5.9	7.29	3.98	5.9
21	6.56	0	4.88	5.22
22	5.9	0	0.02	0
23	0	6.56	0	6.56
24	5.9	7.29	0	7.29
25	6.56	8.1	0	0
26	0	0	5.9	7.29
27	6.56	0	6.56	8.1
28	7.29	0	7.29	9
30	0	0	0	0

B. (10 points) Plot the accumulated reward for the run, i.e. plot the total amount of reward received so far against the number of episodes, and show the greedy policy with respect to the value function.

In order to see the changes between accumulated reward and against the number of episodes, I have selected to plot for the first 110 episodes.

x-coordinate: accumulated reward
y-coordinate: number of episodes



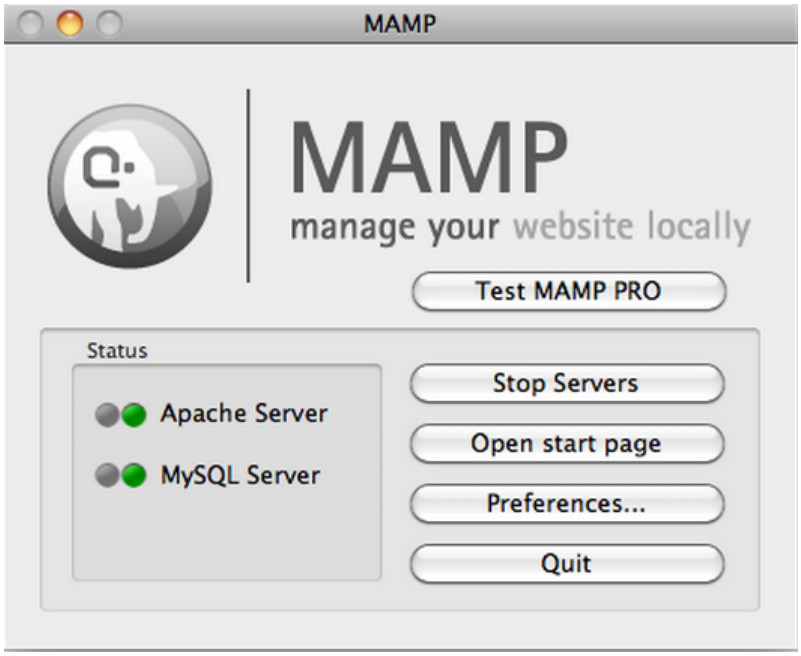
Question3.

(BONUS10points)IfinsteadofmovingN,S,E,W,theagentmoveslikea*knight*inchess (for example, one step North, then two steps East, in one move), how would the value of the states change (run Q-learning with this new action set)?

How to run my codes:

1) Setting up PHP and Apache:

- Download the MAMP disk image from http://www.mamp.info/downloads/releases/MAMP_PRO.zip
- Open the disk image and drag MAMP into your Applications folder.



After successful installation you can launch your local web server. Start MAMP and click on the "Start Servers" button. In the status display on the left, the launch status of the server is displayed. Should you not be logged in as administrator, you are asked at this point for the administrator's password. The web server starts by default on port 8888. This port must be specified when calling the local web page in the browser, e.g.: http://localhost:8888 You can change the port in the "Preferences" button.

Make sure the status display on the left shows the started services (Apache, MySQL). We need only Apache!

2) Run the PHP codes:

Put my PHP codes in the following directory (the Apache in MAMP is pointed to this path by default):

/Applications/MAMP/htdocs/q-learning.php
/Applications/MAMP/htdocs/policy-eva.php

Now, open a web browser and run the codes:

http://localhost:8888/policy-eva.php
http://localhost:8888/q-learning.php