CPSC 502 — Fall 2013
Assignment 4
Solution

Question 1

Solution

(a) There are 4 policies:

- Always exercise.
- Always rest.
- Exercise when fit and rest when unfit.
- Rest when fit and exercise when unfit.

(b) The python program
`http://www.cs.ubc.ca/~poole/cs502/2013/as4/as4sol.py`
prints a solution to this and the next part.

(c) see (b)

(d) As γ increases, the long-term becomes more important, and the optimal policy becomes to always exercise. In the short term, it is better to relax in either state. However, the long term advantage of being fit overcomes the short-term advantage of relaxing when γ is higher.

As γ decreases the long-term becomes less important, and the optimal policy becomes to always rest. In each state, the immediate reward for resting is higher than that of exercising. As γ gets smaller, the short term advantage of relaxing overcomes the long-term advantage of being fit.

Question 2

Solution

(a) The optimal decision tree with one node predicts $Likes = false$ (or not *likes*). It has 6 errors.

Aside: the prediction of *likes* with probability $6/13$ has an error of 6.46.

(b) The optimal prediction is to predict *likes* with probability $6/13$. It has error $6*(7/13)^2+7*(6/13)^2 = 3.23$.

Aside: the prediction of *likes* has an error of 6.

(c) An optimal decision tree of depth 2 is:

```
if comedy then likes=true else likes=false
```

It has 4 errors. At the root are all of the examples $(e_1, \ldots, e_{13})$. Filtered to the $comedy = true$ node are examples $e_2, e_6, e_7, e_{10}, e_{11}, e_{13}$. Filtered to the $comedy = false$ node are $e_1, e_3, e_4, e_5, e_8, e_9, e_{12}$.

Another optimal decision tree of depth 2 is:

```
if lawyers then likes=true else likes=false
```

It has 4 errors. At the root are all of the examples $(e_1, \ldots, e_{13})$. Filtered to the $lawyers = true$ node are $e_2, e_3, e_4, e_8, e_9, e_{10}$. Filtered to the $lawyers = false$ node are $e_1, e_5, e_6, e_7, e_{11}, e_{12}, e_{13}$.

(d) For the sum-of-squares error, the best tree is

```
if comedy then likes=2/7 else likes=4/6
```

The sum of squares error is

$$5(\frac{2}{7})^2 + 2(\frac{5}{7})^2 + 4(\frac{2}{6})^2 + 2(\frac{4}{6})^2$$

(e) The smallest decision tree is

```
if guns then
   {if lawyers then likes=true else likes=false}
 else
   { if comedy  then likes=true else likes=false}
```

(f) One way to find such trees is to do a two-step lookahead. For each property, check a split on that property and then do a split on each leaf before evaluating the split.

Question 3

After running 1000000 steps of the tiny game at `http://artint.info/demos/rl/tGame.html`, Chris was suspicious that Q-learning converges to the correct Q-values where $\alpha_k = \frac{1}{k}$, the empirical average.

Chris then had some hypotheses about which method converges faster (or at all) to the correct Q-values:

(a) $\alpha_k = \frac{1}{k}$

(b) $\alpha_k = \frac{10}{9+k}$

(c) $\alpha_k = 0.1$.

(d) $\alpha_k = 0.1$ for the first 1,000 steps, $\alpha_k = 0.01$ for the next 1,000 steps, $\alpha_k = 0.001$ for the next 1,000 steps, $\alpha_k = 0.0001$ for the next 1,000 steps, and so on.

Either construct the simplest possible RL example you can construct that exhibits slow convergence for the empirical average, or modify the open-source code for the tiny game (α is used in `TGameQController.java`) to investigate how well each of these works.

This is intended to allow an open-ended investigative question. You can get full marks by either answering the specifics of the question or helping Chris in some other way to understand the convergence of the empirical average. Write something interesting with some (theoretical or empirical) evidence.

Solution (a) doesn't seem to converge even after tens of billions of steps. (b) works very well. (c) moves quickly to approximately the correct answer, but varies around that value (d) seems to converges in this example, but isn't guaranteed to (it depends of 1000 is enough to reduce the error by an order of magnitude).