CSE 415 Computer Science and Engineering, University of Washington

Name: _____

Watch the Water!

Droughts have become a big problem in the Northwest. The company wawawa123.com has won a state-government contract for an automated water-watching policy setter. Their system, called the Washington Water Watcher (Wa-Wa-Wa for short) models the state's water system as an MDP. The beta version has just been rolled out, and the company has published all the modeling parameters. However, it's not working correctly, and they need help. What's wrong?

States:

F: Reservoirs are mostly full.

L: Reservoirs are mostly low.

Actions:

Rationing: Ration water for the next 3 months.

No rationing: Do not ration water for the next 3 months.

Transition and reward functions:

Each of the 8 table entries gives a probability (T value) and a reward (R value). For example, from the first entry we can see that T(F, Rationing, F) = 0.8, and R(F, Rationing, F) = 2. There is a pair of rows for each current state *s* and a column for each possible successor *s'*.

s'

s		F	L
	F	Rationing: 0.8, 2	Rationing: 0.2, 2
		No rationing: 0.5, 5	No rationing: 0.5, 4
	L	Rationing: 0.5, 2	Rationing: 0.5, 2
		No rationing: 0.2, 4	No rationing: 0.8, -10

Rationale: When the reservoirs are full, and Wa-Wa-Wa implements rationing, then the reservoirs are likely to stay full. People don't like the rationing very well (reward is only 2). There is still some chance that the reservoirs will go low, but it's only a 0.2 chance.

If you don't implement rationing, people are happy and you get a reward of 5, but there is a 0.5 chance of going low, and there could be some folks even more unhappy if their water goes very low.

When the reservoirs are low, rationing gives you a 0.5 chance of them filling up again and a 0.5 chance of them staying low. People aren't very happy with any rationing, so it's still 2 if the reservoirs go high, and 2 if they stay low.

Without rationing, there is still some slight possibility (0.2) of the reservoirs filling up, and people are happy (but not as happy as when the reservoirs start out full, due to some folks having really low water).

But the worst happens when the reservoirs are low, there is no rationing, then the reservoirs stay low, and during that low-to-low period, many customers have outages. Here the reward is -10.

Draw a state-transition diagram for this Markov Decision Process:

The discount rate gamma is 0.5, reflecting the idea that making people happy today is more important than in the future, especially given the fact that the company doesn't yet have a contract with the state for the next biennium. Note that this is technically, however, an infinite-horizon MDP.

The company's agent is implementing rationing ALL THE TIME. What is it doing wrong?

Use the Value Iterations method to determine a simple approximation of the expected values of each state under the optimal policy that suggests what the problem is.

In particular, assume $V_0(s) = 0$ for each s in {F, L}, and then compute $V_1(s)$ using one step of the Bellman update.

Here is the formula for the Bellman Update operation:

$$V_{k+1}(s) = \max_a \Sigma_{s'} T(s, a, s') [R(s, a, s') + \gamma V_k(s')]$$

$V_0(F) = 0$ $V_0(L) = 0$

$V_1(F) =$

$V_1(L) =$

What policy does your computation suggest would be better?