Value Iteration

Declare:

U', U - current utilities for states, initially $\emptyset$

δ - max change of any utility of any state in an iteration

ϵ - max allowable error of utility in any state. (controls convergence of algorithm)

Requires MDP with:

Actions, $a \in A(s)$: actions in state s

States s

Rewards(s) : rewards in state s

discount γ

Transition Model : $P(s'|a,s)$, aka $T(s,a,s')$

Loop

$U = U'$

$\delta = \emptyset$

for each state in States

$$U'(s) = R(s) + \gamma \max_{a \in A(s)} \sum_{s'} P(s'|a,s)\, U(s')$$

if $\text{abs}(U'(s) - U(s)) > \delta$

$\delta = \text{abs}(U'(s) - U(s))$

Until $\delta < \epsilon(1-\gamma)/\gamma$

$$U_1(2,3) = R(2,3) + \gamma \max\Big[\big(-1\times .8 + .72\times .1 + .1\times 0\big),\ \big(.72\times .8 + -1\times .1 + 0\times .1\big),$$

$$\big(0\times .8 + .1\times -1 + .1\times 0\big),\ \big(.8\times 0 + .1\times 0 + .1\times 0\big)\Big]$$

$$\gamma \max\Big[-.728,\ \cancel{.4284}\ .476,\ -.1,\ 0\Big]$$

$$\cancel{10 + .9\times .4284}$$

$$= 0 + .9\times .476 = .4284$$

		.72	+1
		.4284	-1

(x)

(y)

Fill in values for all states for $U_1(s)$ $\forall s$

~~Restart~~

Utils are closer for states closer to the exit. because fewer steps are required.

Qs (1,2)

Same as (2,3). It's same # of steps? Shouldn't be, there's no fire pit danger.

	?	.72	+1	1
		.4284	-1	2
				3
1	2	3	4	

(x)

(y)

How long does value iteration run?

~~Until~~

Until convergence - when utilites for ~~any~~ all states change less than δ.

Value iteration gives us the utilities, can also store the actions, or go back through the grid and move toward highest utility

Exercises

Given:

10				1
a	b	c	d	e

is reward

Actions: East, West, and Exit (in states a, e only)

Transitions: deterministic

1. For $\gamma = 1$, what is optimal policy? Action for each state

2. For $\gamma = .1$, what is optimal policy?

3. For what γ are E and W equally good when in state d?

$U(s) = R(s) + \gamma \max(\leftarrow, \rightarrow)$

1.

10	←	←	←	1
a	b	c	d	e

exit exit

no discounting, go to highest reward

from d, $\gamma^3 \times 10 = 10$

2. Go west for all and get (from d). Go east from d: $\gamma \times 1 = .1$ (better)

$\gamma^3 \times 10 = .1^3 \times 10 = .01$

from c $\gamma^2 \times 10 = .01 \times 10 = .1$

10	←	←	→	1
a	b	c	d	e

exit exit

$\gamma^2 \cdot 10$ $\gamma \cdot 0$ 0

$.01 \times 10 = .1$

3. $\gamma^3 \times 10 = \gamma \cdot 1 \cdot 0 = \gamma^2 \times 10 = 1$

$\gamma = \sqrt{\frac{1}{10}}$