CSCI-4150 Intro to Artificial Intelligence, Fall 2004

Q learning example

Suppose we are using Q learning for an agent in the following world, similar to the example from the text.

②	④ 1
①	③ −1

States ③ and ④ are terminal states, and the actions (up, down, left, right) are nondeterministic.

Suppose the current Q values are:

$Q(\text{up}, 1) = -0.4$	$Q(\text{up}, 2) = -0.8$
$Q(\text{down}, 1) = 0.5$	$Q(\text{down}, 2) = 0.2$
$Q(\text{right}, 1) = 0.8$	$Q(\text{right}, 2) = 1$
$Q(\text{left}, 1) = -0.1$	$Q(\text{left}, 2) = 0.6$

~~Since states 3 and 4 are terminal states, we will assume~~

~~that we~~ Since states 3 and 4 are terminal states from which we will see no transitions, we will assume $Q(a,3) = Q(a,4) = 0$ (for any action a)

A) Compute a policy for this world.

B) Suppose you see the following transitions:

① From state 1 to state 3 under action "up" with reward −1

② From state 1 to state 2 under action "up" with reward −0.1

③ From state 2 to state 4 under action "down" with reward +1

Do Q learning on each of these transitions.

Use $\alpha = 0.1$ (and $\gamma = 1$)