Q-Learning Algorithm Explained with Example

9/22/17

Nathan Metze

https://cdn-images-1.medium.com/max/1600/1*2RjuuBUk5ArFAGjaU4jiwA.png

s = represents current state

a = action taken from current state

s’ = represents the state resulting from the action

r = reward you get for taking the action

y = the discount factor

EXAMPLE:



With initial Q-table:



U – up

D – down

L – left

R – right

N – no action

With initial R-table:



E = NULL

ALGORITHM:

Algorithm:

1. Initialize Q-matrix by all zeros. Set value for ‘γ’. Fill rewards matrix.
2. For each episode. Select a random starting state (here we will restrict our starting state to state-1).
3. Select one among all possible actions for the current state (S).
4. Travel to the next state (S’) as a result of that action (a).
5. For all possible actions from the state (S’) select the one with the highest Q value.
6. Update Q-table using eqn.1.
7. Set the next state as the current state.
8. If goal state reached then end.

Example: Let’s say we start with state 1. We can go either D or R. Say, we chose D. Then we will reach 3 (the snake pit). So, then we can go either U or R. So, taking γ = 0.8, we have:

Q(1, D) = R(1,D) + γ\*[max(Q(3,U) & Q(3,R))]

Q(1, D) = -10 + 0.8\*0 = -10

Here, max(Q(3, U) & Q(3,R)) = 0 as Q matrix not yet updated. -10 is for stepping on the snakes. So, new Q-table looks like: