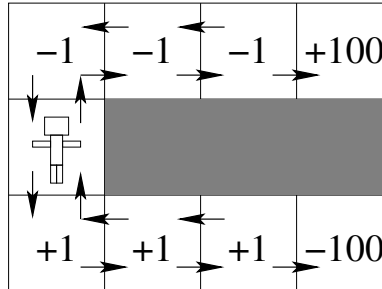


Homework 10

CS 6243 – Spring 2005
Tom Bylander, Instructor

assigned April 14, 2005
due April 21, 2005

(100 pts.) Consider the following environment.



The +100 and -100 states are terminal states. The robot always starts at the left center state and must choose which neighboring state to move to each time step. Moves are deterministic. The reward for moving to a state are shown.

1. (20 pts.) What is the value of a policy that moves to and stays within the +1 states? Assume a discount of 0.9.
2. (20 pts.) What is the value of a policy that moves directly to the +100 state? Assume a discount of 0.9. This should have a much higher value than the previous policy.
3. (20 pts.) For each of the +1 states, determine the value of a policy that goes directly to the +100 state.
4. (20 pts.) Why will the optimal policy be hard to learn? Hint: what are the chances that an ϵ -greedy exploration strategy will reach the +100 state?
5. (20 pts.) Suppose the reward of the +100 state is lowered to x . What values of x make the “move to and stay within the +1 states” a better policy? Again assume a discount of 0.9.
6. (100 pts., shared extra credit) Suppose the moves are nondeterministic. Specifically, suppose there is a 10% chance that the robot moves the opposite direction. Determine the optimal policy and the value of each state. Note: this would be like an ϵ -greedy strategy for the deterministic version of the problem with $\epsilon = 0.2$.