CS188: Artificial Intelligence, Fall 2008
Section #5 – Golf as an MDP

In this exercise we will formulate golf as an MDP as follows:

- State Space : {Tee, Fairway, Sand, Green}
- Actions : {Conservative shot, Power shot}
- Initial State : Tee
- Terminal State : Green
- Transition model : (note that any successor state not on this list has a transition probability 0, and "Conservative" stands for "Conservative shot")

s	a	s'	$T(s, a, s')$
Tee	Conservative	Fairway	0.9
Tee	Conservative	Sand	0.1
Tee	Power shot	Green	0.5
Tee	Power shot	Sand	0.5
Fairway	Conservative	Green	0.8
Fairway	Conservative	Sand	0.2
Sand	Conservative	Green	1.0

- Rewards:

  (note: $R(\cdot, \cdot, s)$ means that the reward is received for transitioning to state s, regardless of the action taken or previous state)

s	$R(\cdot, \cdot, s)$
Fairway	-1
Sand	-2
Green	3

1

a) Draw a state graph defining this MDP problem. A state graph shows the states as nodes and has the actions as arcs labeled with T and R values. Remember, in a state graph no states repeat.

2 Conservative Policy

Consider the policy of always taking the "Conservative shot".

a) To help with (2b), draw the MDP search tree corresponding to this problem for the initial state. Be clear about which nodes are the MAX nodes and which are the chance nodes. Since this is for a specific policy (always taking the "Conservative shot"), the MAX nodes will only contain a single child. (You do not need to include nodes that are impossible to reach under the specified transition probabilities.)

b) What is the utility of the initial state under the "Conservative shot" policy from part (a)? (Assume the discounting factor $\gamma = 1$.)

3 Optimal Policy

a) Complete the table below containing estimates of the utility of each state using Value Iteration for the first 3 iterations. Assume we start with all utilities set to 0 and $\gamma = 1$.

State	$iter. = 0$	$iter. = 1$	$iter. = 2$	$iter. = 3$
Tee	0			
Fairway	0			
Sand	0			
Green	0			

b) Assuming that the utilities computed in the previous question are optimal, what is the optimal policy?