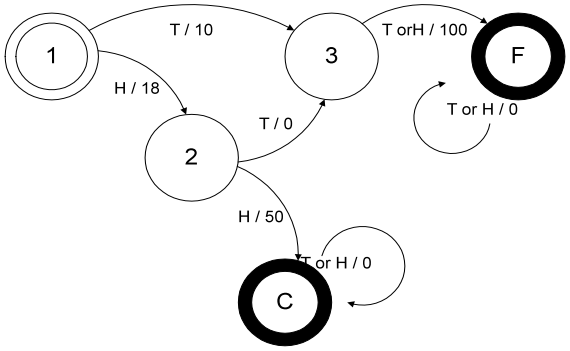
Massachusetts Institute of Technology

16.410-13 Principles of Automated Reasoning and Decision Making

Problem Set #11 Do not Turn In

Problem 1 –MDPs: Tortoise and Hare

The following question is taken from the 2004 final. We all know, as the story goes, that the Tortoise beat the Hare to the finish line. The Tortoise was slow, but extremely focused on the finish line, while the Hare was fast, but easily distracted. Although the Tortoise crossed the finish first, who really gained the greatest reward, the Tortoise or the Hare? It's a matter of perspective. To resolve this age old question, we frame the race as an MDP, solve for the optimal policy, and use this policy to determine once and for all whose path is best, the Tortoise or the Hare.



State	Action	Next State	Reward
1	T	**3**	10
1	H	**2**	18
2	H	**C**	50
2	T	**3**	0
3	T or H	**F**	100
C	T or H	**C**	0
F	T or H	**F**	0

We model the race with the above MDP. The race starts at 1, and finishes at F. 2 and 3 denote intermediate check points along the race course, while C denotes a Cabbage patch, which is very enticing to the Hare. Actions are T and H. T denotes actions focused towards the finish line, while H denotes an action that grabs the greatest immediate reward. The tortoise's sequence <T, T> is the shortest path to the finish line. The hare's sequence <H,H> is the direct path to the cabbage patch, with rewards along the way. <H, T, T> represents a mixed strategy, balancing immediate and long term reward.

Part A. Value Function and Policy for Tortoise Discount

Use value iteration to determine the infinite horizon, discounted lifetime reward for states 1-3. For this part the Tortoise provides us with a discount factor, which is $\gamma = 0.9$. **Fill in the missing entries in the table below**. On the left, V^i denotes the value function at iteration i. **Stop** iterating as soon as $\mathbf{V^i}$ **converges**. Next **fill in** the actions for the **optimal policy on the right**.

State	V^0	V^1	V^2	V^3	V^4	V^5	π^*	State	Action
1	0							1	
2	0							2	
3	0							3	
F	0								
C	0								

Derive your answers below.

Part B. Value Function and Policy for Hare Discount

The Hare complains that the model is unrealistic, "one should always live in the moment" he says. Instead he gives you a new discount factor of $\gamma = 0.1$. To give the Hare the benefit of the doubt, recompute V and π^* with this new γ. Fill in the following table, according to the same directions as Part A.

State	V^0	V^1	V^2	V^3	V^4	V^5	π^*	State	Action
1	0							1	
2	0							2	
3	0							3	
F	0								
C	0								

Derive your answers below.

Part C. A γ That Splits Hares

In order to get the Hare to head towards the finish line, we will need to perform the difficult task of training the Hare to discount the future less. What is the minimum value of the Hare's discount factor, so that the Hare heads to the finish line? That is, compute the greatest lower bound on γ such that the Hare does not head to the cabbage patch.

Answer: $\_\_ \leq \gamma$

Prove that your bound is correct:

Problem 4: Time (5 pts)

Please let us know the amount of time it took you to complete this problem set. Please separate the amount of time for the written and for the coding components.

16.410 / 16.413 Principles of Autonomy and Decision Making

Fall 2010