Tutorial 5

Markov Decision Process

Consider a robot that is moving in an environment. The goal of the robot is to move from an initial point to a destination point as fast as possible. However, the robot has the limitation that if it moves fast, its engine can overheat and stop the robot from moving. The robot can move with two different speeds: *slow* and *fast*. If it moves fast, it gets a reward of 10; if it moves slowly, it gets a reward of 4. We can model this problem as an MDP by having three states: *cool*, *warm*, and *off*. The transitions are shown in below. Assume that the discount factor is 0.9 and also assume that when we reach the state *off*, we remain there without getting any reward.

$\boldsymbol{s}$	a	$\boldsymbol{s'}$	$\boldsymbol{P(s'\|a,s)}$
Cool	Slow	Cool	1
Cool	Fast	Cool	1/4
Cool	Fast	Warm	3/4
Warm	Slow	Cool	1/2
Warm	Slow	Warm	1/2
Warm	Fast	Warm	3/4
Warm	Fast	Off	1/4

1) Consider the conservative policy when the robot always moves slowly. What is the value of $v_*(Cool)$ under the conservative policy? Remember that $v_*(Cool)$ is the expected discounted sum of rewards when starting at state s.

2) What is the optimal policy for each state?