MDPs Study Guide (AIMA 17)

Artificial Intelligence

1 Markov Decision Processes

1. What are the components of a Markov Decision Process?

Solution: A Markov decision process (MDP) is a 4-tuple $(S, A, Pr(s' \mid s, a), R(s))$, where

- S is a set of states,
- A, or $Action(s)$ is a set of actions, and
- $Pr(s' \mid s, a)$ is a transition function giving the probability that executing action a in state s will result in s'. Many authors use $T(s, a, s')$
- $R(s, a, s')$ is the reward the world provides to an agent for arriving in state s' after executing action a in state s, bounded by $\pm R_{max}$. Many authors use $R(s')$, which is easier to think about – the reward for arriving in state s' regardless of the s, a pair in the previous time step.

Some definitions of MDPs include an initialization function, $I(s)$, which specifies the probability the the agent will start in some state $s \in S$, others specify a particular state s_0 from S as the start state.

2. What is the role of γ in the calculation of infinite horizon returns, $G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$?

Solution: To discount the value of future rewards.

2 Policies

1. In the context of Markov Decision Processes, what is a policy?

   Solution: A policy is a function that returns a recommended action for every state, denoted $\pi(s)$.

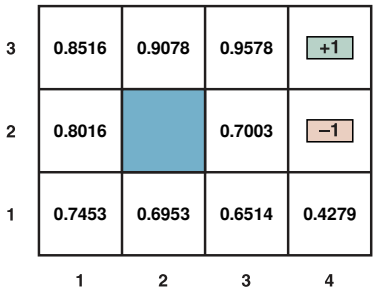
2. In the context of Markov Decision Processes, what is a stochastic policy?

   Solution: A stochastic policy is a probability distribution over actions conditioned on the state, $\pi(a \mid s)$.

3. How many optimal policies are there in an MDP?

   Solution: At least one

4. Given the following optimal state values:



   What is the deterministic optimal policy, $\pi(s)$?

$\pi((1,3)) = Right$	$\pi((2,3)) = Right$	$\pi((3,3)) = Right$	$\pi((4,3)) = Noop$
$\pi((1,2)) = Up$		$\pi((3,2)) = Up$	$\pi((4,2)) = Noop$
$\pi((1,1)) = Up$	$\pi((2,1)) = Left$	$\pi((3,1)) = Up$	$\pi((4,1)) = Left$