UCLRL Lecture 2 Notes

August 5, 2018

1 Markov Decision Processes

1.1 Markov Reward Processes

Definition 1. *A Mariv Process is a tuple* $\langle \mathcal{S}, \mathcal{P} \rangle$

For a Markov state s and successor state s', the state transition probability is defined by

$$\mathcal{P}_{ss'} = \mathbb{P}\left[S_{t+1} = s' | S_t = s\right]$$

and we can characterize the transition from all states by the transition matrix $\mathcal{P}$ where

$$\mathcal{P} = \begin{bmatrix} \mathcal{P}_{11} & \dots & \mathcal{P}_{1n} \\ \vdots & & \\ \mathcal{P}_{n1} & \dots & \mathcal{P}_{nn} \end{bmatrix}$$

Definition 2. *A Markov reward process is a Markov process but with a tuple* $\langle \mathcal{S}, \mathcal{P}, \mathcal{R}, \gamma \rangle$ *such that*

- $\mathcal{R}$ *is a reward function such that* $\mathcal{R}_s = \mathbb{E}\left[R_{t+1} | S_t = s\right]$
- γ *is a discount factor with* $\gamma \in [0, 1]$

Definition 3. *The* ***return*** G_t *is the total discounted reward from time-step* t

$$G_t = R_{t+1} + \gamma R_{t+2} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

Definition 4 (State Value Function)**.** *The state value function* $v(s)$ *of an MRP is the expected return starting from state* s

$$v(s) = \mathbb{E}\left[G_t | S_t = s\right]$$

The value function be decomposed into

- immediate reward R_{t+1}
- discounted value of successor state $\gamma v(S_{t+1}$

$$\begin{aligned} v(s) &= \mathbb{E}\left[G_t | S_t = s\right] \\ &= \mathbb{E}\left[R_{t+1} + \gamma G_{t+1}\right] \\ &= \mathbb{E}\left[R_{t+1} + \gamma v(S_{t+1}\right] \end{aligned}$$

Definition 5 (Bellman Equation for MRP)**.**

$$v(s) = \mathbb{E}\left[R_{t+1} + \gamma v(S_{t+1}|S_t = s\right]$$

which can be rewritten as

$$v(s) = \mathcal{R}_s + \gamma \sum_{s' \in \mathcal{S}} P_{ss'} v(s')$$

or in matrix notation

$$v = \mathcal{R} + \gamma \mathcal{P} v$$

Bellman equation can be solved direcly as

$$\begin{aligned} v &= \mathcal{R} + \gamma \mathcal{P} v \\ &= (I - \gamma \mathcal{P})^{-1} \mathcal{R} \end{aligned}$$