CMPUT 655: RL–1 Lecture 3

14th September 2020

Contents

We will talk about

- Small recap about what we did last time about probability
- And discuss these a little bit in the context of MDPs

Reminder about Some Probability Rules

- Conditional probability (also chain rule)

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

an outcome

die #

$$P(A,B,C) = \underbrace{P(A,B|C)}\, P(C)$$

$$P(X = 1 | Y = \{H\})$$

- Marginilization over a joint probability distribution

$$P(X = x) = \sum_{y \in \mathcal{Y}} \underbrace{P(X = x, Y = y)}.$$

Reminder about Some Probability Rules (contd.)

- Conditional Expectation

$\mathbb{E}[X] = \sum_{x} x \, p(x)$

$$\mathbb{E}_X[X|Y = y] = \sum_{x \in \mathcal{X}} x \cdot p(X = x|Y = y).$$

- ~~Conditional Expectation and~~ Law of Total Expectation

$$\mathbb{E}_X[X|Y = y] = \mathbb{E}_Y\left[\mathbb{E}_X[X|Y = y]\right].$$

RL Book Notation

- Mathcal (fancy) Symbols:

s_0, s_1, s_2

$\mathcal{S} \rightarrow$ {behind a car, speeding, dropping, ...}

- Capital Symbols:

$S_0 \rightarrow$ r.v. stores the state at $(t=0)$

A_0 — action at $t=0$

- Small Symbols:

s_0 = behind a car

A_0 = steer left $\rightarrow a_1$

Example MDP

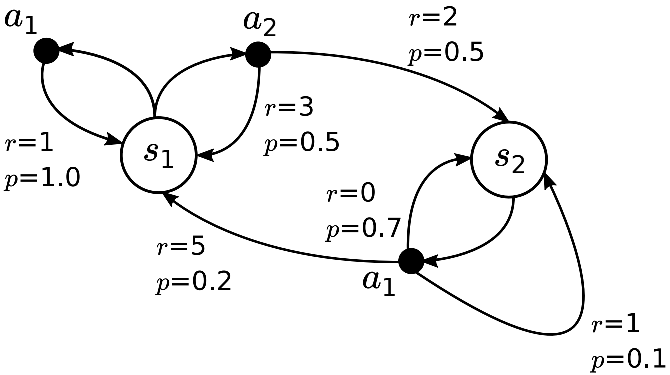

Figure: MDP $M = (\mathcal{S}, \mathcal{A}, \mathcal{R}, P, \gamma)$.

$\{s_1, s_2\}$

Agent's Interaction and the Trajectory ($\sim$ stream of experience)

s_1, a_1, 1, s_1, a_1, 1, ...

$\| \quad \| \quad \| \quad \| \quad \|$

S_0 A_0 R_1 S_1 A_1 ...

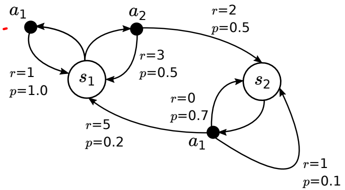

Figure: Trajectory $\tau_t = (S_0, A_0, R_1, S_1, A_1, R_2, S_2, \ldots, S_T)$.

Let us calculate the probability of seeing this trajectory:

$$P(S_0, A_0, R_1, S_1, A_1, \ldots) = P(S_0) \times P(A_0 R_1 S_1, \ldots \mid S_0)$$

$$= P(S_0)\, \pi(A_0 \mid S_0)\, P(S_1 R_1 \mid A_0 S_0) \ldots$$

Different Transition Probabilities (Section 3.1 RL Book[1])

- $p(s', r|s, a) \doteq \Pr\left(S_{t+1} = s', R_{t+1} = r \middle| S_t = s, A_t = a\right)$

  marginal joint conditional

- $p(s'|s, a) \doteq \Pr(S_{t+1} = s' \mid S_t = s, A_t = a)$

  $= \sum_{r \in \mathcal{R}} p(s', r|s, a)$

- (Dependent on the policy π)
- $p(s'|s) =$

  $p(s', r|s, a) \quad \pi(a|s)$

  V Q

  $p(s', r|s, a) \xrightarrow{r} p(s'|s, a)$

  $p(s', a|s) = p(s'|s, a) \times \pi(a|s)$

  $p(s'|s) = \sum_a p(s', a|s) = \sum_a p(s'|s, a)\pi(a|s)$

- $p(s', a'|s, a) = \pi(a'|s')\, p(s'|s, a)$

$p(s_2|s_2, a_1) = 0.1 + 0.7 = 0.8$

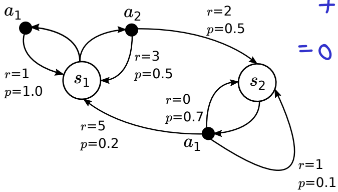

Figure: MDP

[1] http://incompleteideas.net/book/the-book-2nd.html

Different Reward Functions (Section 3.1 RL Book[2])

- The random variable R_t

- Reward function

$p(r|s,a)$

$$r(s,a) \doteq \mathbb{E}\left[R_t \middle| S_{t-1} = s, A_{t-1} = a\right] \quad \leftarrow \text{cond. Expectation}$$

$$= \sum_r r \cdot p(R_t = r \mid S_{t-1} = s, A_{t-1} = a)$$

$\sum_{s'} p(s',r|s,a)$

$$= \sum_r r \sum_{s'} p(s',r|s,a)$$

- $r(s,a,s') \doteq \mathbb{E}\left[R_t \mid S_t = s', S_{t-1} = s, A_{t-1} = a\right]$

$$= \sum_r r \cdot p(r|s',a,s)$$

$$p(r|s',a,s) = \frac{p(s',r|sa)}{p(s'|s,a)}$$

- $r_\pi(s) \doteq \mathbb{E}_\pi[R_t|S_t = s] =$

cond. bayes. margin.

Mossi π

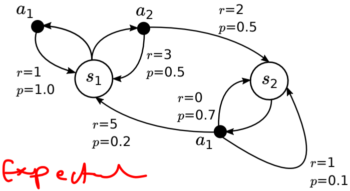

Figure: MDP

[2]http://incompleteideas.net/book/the-book-2nd.html

Bellman Equation ($V^{\pi} \to V^{*}$)

Read from back

S

Bellman Equation ($V \to V$) (using $r(s, a)$)

What does $\mathbb{E}_\pi$ mean?

$\pi(a|s) = P(A_t = a | S_t = s)$

$$V_\pi(s) = \mathbb{E}_\pi \left[R_1 + \gamma R_2 + \gamma^2 R_3 + \cdots \,\middle|\, S_0 = s \right]$$

$$= \sum_{a_1 \in \mathcal{A}(s)} \pi(a_1|s) \sum_{s_1, r_1} p(s_1, r_1 | s, a_1) \left[r_1 + \gamma \sum_{a_1} \pi(a_1|s_1) \sum_{s_2 r_2} p(s_2, r_2 | s_1, a_1) \left[r_2 + \gamma \cdots \right.\right.$$

$S_2\ R_2$

a_0 is an instantiation at $t = 0$ (A_0)

$a_i \in \mathcal{A}(s_i)$ all the actions at $t = i$

x	y	z
a_0	a_1	a_2

$$\mathbb{E}_\pi \cong \mathbb{E}_{A_0 \sim \pi(\cdot|s),\ S_1 R_1 \sim P(\cdot,\cdot|s, A_0);\ A_1 \sim \pi(\cdot|s_1) \cdots}$$

$$s \xrightarrow[R_1]{A_0 \sim \pi} S_1 \xrightarrow{A_1 \sim \pi} \cdots$$

$$\mathbb{E}_X[X] = \sum_{x \in \mathcal{X}} x \cdot p(X = x)$$

Bellman Equation ($V \to V$) (more thoughts)

Summary