Lec-4, IT567, 24-25

Agent & environment interact at each of a seq. of discrete time steps, $t = 0, 1, 2, 3 \ldots$

At time t, agent in a state S_t ($\in \mathcal{S}$), takes an action $A_t \in \mathcal{A}(S_t)$

$\mathcal{A}(S_t) \rightarrow$ set of possible actions, an agent can take in state S_t.

We can say $\mathcal{A}(S_t) = \mathcal{A}(S)$

$\Downarrow$

action set is the same in all states.

$\mathcal{S} = \{s_1, s_2, s_3, s_4\}$

$S_0 = s_1 \quad S_4 = s_1$

$S_1 = s_3$

$S_2 = s_4$

$S_3 = s_2$

$t=$	0	1	2	3	4
	s_1	s_3	s_4	s_2	s_1

$\mathcal{A}(s_1), \mathcal{A}(s_2), \mathcal{A}(s_3)$

$\mathcal{A}(s_4), \mathcal{A}(s_5)$

Agent receives reward, $R_{t+1} \in \mathcal{R} \in \mathbb{R}$ & finds itself in a new state S_{t+1}.

$\mathcal{R} = \{1, -1, 0.5, 7.5 \ldots\}$

Finite MDPs:- The sets $\mathcal{S}$, $\mathcal{A}$ & $\mathcal{R}$ all have a finite number of elements.

now, R_t & S_t are random variables belonging to the set. Hence we define

MDP:

$$p(s', r \mid s, a) \triangleq \Pr\{S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a\}$$

$$\forall \quad s', s \in \mathcal{S} \quad \& \; a \in A(s) = \mathcal{A}$$
$$r \in \mathcal{R}$$

The function $p(\cdot\cdot \mid \cdot\cdot)$ defines the dynamics of the MDP.

$p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \longrightarrow [0, 1]$

is an ordinary deterministic function of 4 arguments.

$A_1 = \{1, 2, 3\}$

$B_1 = \{4, 5\}$

$A_1 \times B_1 = ?$

$p(s', r \mid s, a) \Rightarrow$ RVs. R_t & S_t have well defined discrete prob. distributions dependent only on the preceeding state & action — Markov prob.

now, $$\sum_{s' \in \mathcal{S}} \sum_{r \in \mathcal{R}} p(s', r \mid s, a) = 1 \quad \forall \; s \in \mathcal{S}, a \in \mathcal{A}(s)$$

$\because$ conditional prob. distribution is also a prob. distribution

We can compute state-transition probabilities as

$$p(s' \mid s, a) \triangleq Pr\{S_t = s' \mid S_{t-1} = s, A_{t-1} = a\} = \sum_{r \in \mathcal{R}} p(s', r \mid s, a)$$

given $p(s', r \mid s, a)$

$\mathcal{S} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$

$E[X \mid Y = y]$

$\int_x x \, p(x \mid y)$

$$r(s, a) \triangleq E[R_t \mid S_{t-1} = s, A_{t-1} = a]$$

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

$$\sum_{s' \in \mathcal{S}} p(s', r \mid s, a)$$

expected reward for state-action pairs as a two argument function $r: \mathcal{S} \times \mathcal{A} \to \mathbb{R}$

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

$$= \sum_{r \in R} r \cdot \frac{p(s', r \mid s, a)}{p(s' \mid s, a)}$$

expected rewards for state-action-next state triples as a 3-argument function

$$r : S \times A \times S \to \mathbb{R}$$

$$p(x \mid y_1, y_2, y_3) \overset{?}{=} p(x, y_1 \mid y_2, y_3)$$

$$p(x \mid y) = \frac{p(x, y)}{p(y)} = \text{Use Bayes' theorem.}$$