Dynamic Programming

Dynamic Programming; Category: 최적화이론, 알고리즘 ...

큰 문제를 작은 문제로 나누어 푸는 것

ex) 수학적 귀납법

$$F(0) \land \left(\forall n, F(n) \Rightarrow F(n+1)\right) \Rightarrow \forall n, F(n+1)$$

Def. DP := 목적함수(Optimal Object, $W(x)$)를 최대화/최소화 하는 관계식을 찾아내는 것

$$W_n = \sup_{W_n}\{\,f(W_{n+1}, a_{n+1})\}$$: top-down

$$W_n = \max\{\,f(W_{n-1}, a_{n-1})\}$$: bottom-up

Find f that maximizes W

Def. Plant eq.

- $x_t \in X$: state at time t
- $a_t \in A_t$: action at time t

$$f_t: X \times A_t \to X \ \ that \ \ is \ \ f_t(x_t, a_t) = x_{t+1}$$

Def. DP

- r_t: reward at time t
- $\tilde{a} \in (a_0, \dots a_T)$: path of actions through time

Maximize sum of rewards

$$R(\tilde{a}) = \sum_{t=0}^{T} r_t(x_t, a_t)$$

$$R_\tau(\tilde{a}_\tau) = \sum_{t=\tau}^{T} r_t(x_t, a_t) = r_t(x_t, a_t) + R_{\tau+1}(x_{\tau+1}, a_{\tau+1})$$

Maximize

$$W_\tau(\tilde{a}_\tau) = \max_{\tilde{a}_\tau} R_\tau(\tilde{a}_\tau)$$

Def. Bellman eq.

- $W_T(x) = r_T(x)$
- $W_t(x_t) = \sup_{a_t \in A_t} \{r_t(x_t, a_t) + W_{t+1}(x_{t+1})\}$

where $x_t \in X \ \wedge \ x_{t+1} = f_t(x_t, a_t)$

Maximize

$$W_\tau(\tilde{a}_\tau) = \max_{\tilde{a}_\tau} R_\tau(\tilde{a}_\tau)$$

$$\begin{aligned} W_t(x_t) &= \max_{\tilde{a}_t} R_t(\tilde{a}_t) \\ &= \max_{a_t} \max_{\tilde{a}_{t+1}} \{r_t(x_t, a_t) + R_{t+1}(x_{t+1}, a_{t+1})\} \\ &= \max_{a_t} r_t(x_t, a_t) + \max_{\tilde{a}_{t+1}} R_{t+1}(x_{t+1}, a_{t+1}) \\ &= \max_{a_t} r_t(x_t, a_t) + W_{t+1}(x_{t+1}) \end{aligned}$$

Problem.

Plant eq.

$$x_{t+1} = x_t + rx_t(1 - a_t)$$

- $x_0 = x,\ r$: constant
- $0 \le a_t \le 1$: variable

Total Rewards

$$W_0 = ra_0 + ra_1 + \cdots ra_{T-1} = \sum_{t=0}^{T-1} ra_t = R(\tilde{a})$$

Maximize

$$W_\tau(\tilde{a}_\tau) = \max_{\tilde{a}_\tau} R_\tau(\tilde{a}_\tau)$$

Partial Total Rewards

$$W_\tau \coloneqq \sum_{t=t}^{T-1} ra_t = R_\tau(\tilde{a}_\tau) = r_t(x_t, a_t) + R_{\tau+1}(x_{\tau+1}, a_{\tau+1})$$

- $t = T - 1,$

$$W_{T-1} = \max\{rx_{T-1}a_{T-1}\}, \text{ so } a_{T-1} = 1$$

- $t = T - 2,$

$$\begin{aligned} W_{T-2} &= \max_{0 \le a_{T-2} \le 1}\{rx_{T-1}a_{T-1} + W_{T-1}(X_{T-1})\} \\ &= \max\{rx_{T-1}a_{T-1} + r[x_{T-2} + rx_{T-2}(1 - a_{T-2})]\} \\ &= rx_{T-2}\max\{(1 + r) + (1 - r)a_{T-2}\} \\ &= rx_{T-2}\max(1 + r, 2) \end{aligned}$$

Partial Total Rewards

$$W_\tau \coloneqq \sum_{t=t}^{T-1} ra_t = R_\tau(\tilde{a}_\tau) = r_t(x_t, a_t) + R_{\tau+1}(x_{\tau+1}, a_{\tau+1})$$

- $t = T - \mathrm{S},$

Assume, $W_{T-S+1}(x_{T-S+1}) = rx_{T-S+1} \cdot \rho_{T-S+1}$

$$\begin{aligned} W_{T-S}(x_{T-S}) &= \max_{a_{T-S}}\{rx_{T-S}a_{T-S} + rx_{T-S+1}\rho_{T-S+1}\} \\ &= \max\{rx_{T-S}a_{T-S} + r\rho_{T-S+1}[x_{T-S} + rx_{T-S}(1 - a_{T-S})]\} \\ &= rx_{T-S}\max\{(1+r)\rho_{T-S+1} + (1 - r\rho_{T-S+1})a_{T-S}\} \\ &= rx_{T-S}\max((1+r)\rho_{T-S+1}, 1 + \rho_{T-S+1}) \end{aligned}$$

- $Thus,$

$$\therefore \rho_{T-S} = \max\big((1+r)\rho_{T-S+1}, 1 + \rho_{T-S+1}\big)$$