Class 7

Shikhar Saxena

January 27, 2023

Contents

Deterministic Dynamic Programming

Definition 1.

$$V(s_0) := \max_{\pi} V^{\pi}(s_0)$$
$$\textit{where } V^{\pi}(s_0) := \sum_{t=0}^{T-1} r_t(s_t, \pi_t) + r_T(s_T).$$
$$\textit{and } \pi^* := arg \max_{\pi} V^{\pi}(s_0)$$

Let $\pi_t := (\pi_t, \dots, \pi_{T-1})$

Define $V_t^{\pi_t}(s_t) = \sum_{u=t}^{T-1} r(s_u, \pi_u) + r_T(s_T)$

Then, $V_T^{\pi_T}(s) = r_T(s)$ and $V_0^{\pi}(s) = V(s_0)$

Principle of Optimality

$$V_t(s) = \max_{a \in \mathcal{A}} \{r(s,a) + V_{t+1}(s')\} \text{ where } s' = f(s,a)$$

and for $t = T-1, \dots, 0$ set

MISSED?

Blackwell's Principle of Irrelavent Information

Tells that having more information (history) doesn't really help. If the information is not relavent then nah.

Theorem 1. *State space S and another random variable Y (irrelavent info). We want to minimize $E[c(S, A, W)]$ so we can take a policy $\pi : S \times Y \to A$.*

But if W and Y are conditionally independent given S then we can just resort to taking a policy $\pi : S \to A$ and the π^* in this domain will perform better than the one we choose.

MDP

S_t and A_t: Capital Notation because we don't know the deterministic dynamics of the system (here).

Types of Policies

- $\pi_t : (s_t, t) \to \mathcal{A}$, its a Markovian, **non-stationary** and deterministic policy.
- $\pi_t : s \to \mathcal{A}$, its a Markovian, stationary and deterministic policy.

Let Π^K be the space of policies.