Class 8

Shikhar Saxena

January 31, 2023

Contents

MDPs

- Times $t = 0, 1, \ldots, T$
- $\mathcal{S}$: State Space
- $\mathcal{A}$: Action Space
- $S_{t+1} = f_t(S_t, A_t, W_t)$

$$V_t^{\pi_t}(s_t) := E\left[V_{t+1}^{\pi_{t+1}}(s')\right] + r_t(s_t, \pi_t)$$

Here Expectation is not over revenue as we have assumed it to be deterministic (over the state). (This is when MDP is a Markov Reward Process).

$$Q_t(s, a) = r_t(s, a) + E_{s,a}\left[V_{t+1}(f_t(s, a, W_t))\right]$$

Example

Two state MDP

Notation: `(reward, probability)`

- State S_1
  - ⋆ Action $a_{1,1}$
    - $\{5, .5\}$: come back to S_1
    - $\{5, .5\}$: go to S_2
  - ⋆ Action $a_{1,2}$
    - $\{10, 1\}$: goto S_2
- State S_2
  - ⋆ Action $a_{2,1}$
    - $\{-1, 1\}$: goto S_1

Write the Bellman Optimality Equation and find the optimal policy.

(Assume the N for the finite horizon problem and solve for that).