Notes CME241

Raphael Abbou

January 12, 2020

1 Lecture 1-2: Markov Decision Processes

1.1 Definitions

A state S_t represent all the information at time t.

$$\mathbb{P}(S_{t+1}|S_t) = \mathbb{P}(S_{t+1}|S_1, ...S_t)$$

We define the state transition matrix by $\mathcal{P}_{ss'} = \mathbb{P}(S_{t+1} = s'|S_t = s)$

Definition 1.1. *Markov Process (MP): $< \mathcal{S}, \mathcal{P} >$*
Markov Reward Process (MRP): $< \mathcal{S}, \mathcal{P}, \mathcal{R}, \gamma >$
Markov Decision Process (MDP): $< \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma >$

For a MRP:
- $\mathcal{R}$ is a reward function: $\mathcal{R}_s = \mathbb{E}[R_{t+1}|S_t = s]$
- $\gamma \in [0, 1]$ is the discount factor
- $G_t = \sum_k \gamma^k T_{t+k+1}$ is the return

For a MDP:
- $\mathcal{A}$ is the set of actions
- The probabilities of transition depends on the action: $\mathbb{P}^a_{ss'}, \mathcal{R}^a_s$

Definition 1.2. *For MDP, a policy π is a distribution over actions given a state:*

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

Definition 1.3. *For a MDP, we define:*
- *State-value function:*

$$v_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s]$$

- *Action-value function:*

$$q_\pi(s, a) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a]$$

A MDP with a policy π gives:
- A MP $< S, \mathcal{P}^p i >$
- A MRP $< S, \mathcal{P}^p i, \mathcal{R}^p i >$

1.2 Bellman's Equations

For a MRP:

$$v(s) = \mathcal{R}_s + \gamma \sum_{s' \ inS} \mathcal{P}_{ss'} v(s')$$

Or in its matrix form:

$$v = \mathcal{R} + \gamma \mathcal{P} v, v \ in \mathbb{R}^n$$

The resolution of this equation takes $O(n^3)$ (matrix inversion). For large MDP, it can be solved by DP, MC evaluation, or Temporal Difference learning.

For a MDP:

$$v_\pi = \mathcal{R}^\pi + \gamma \mathcal{P}^\pi v_\pi$$

We are looking for:
- optimal state value $v^*(s) = \max_\pi v_\pi(s)$
- optimal action value $q_*(s, a) = \max_\pi q_\pi(s, a)$

How to find the associated optimal policy π^* ? By $a = argmax_{a \in \mathcal{A}} q^*(s, a)$. But we have to know $q^*(s, a)$...

⇒ Obtain it recursively by Bellman Equations

Definition 1.4. *Bellman's Optimality Equations:*

$$v^*(s) = \max_a \mathcal{R}_s^a + \sum_{s' \ inS} \mathcal{P}_{ss'}^a v^*(s')$$
$$q^*(s, a) = \mathcal{R}_s^a + \sum_{s' \in S} \mathcal{P}_{ss'}^a \max_{a'} q^*(s', a')$$

Iterative solution methods: Value Iteration, Policy Iteration, Q-learning, SARSA