

Dynamic Programming & Monte Carlo Methods

DP: mathematical optimization method & computer programming method that refers to simplifying a problem by breaking it down to simpler sub-problems in a recursive manner.

Policy Evaluation & Policy Iteration

$V_{\pi}(s)$ - Expected reward starting at s & following π

- Quantifies how good it is to be in s using π

$Q_{\pi}(s, a)$ - Expected reward starting at s & following policy π , taking action a

π^* - optimal policy with highest long term reward

Policy Iteration

- have initial policy π_0 & value fn V_{π_0}
- Evaluate π_i & find V_{π_i}
- Improve V_{π_0} using greedy strategy
- Converge to π^*

$\pi_0 \xrightarrow{-\epsilon} V_{\pi_0} \xrightarrow{-1} \pi_1 \xrightarrow{-\epsilon} V_{\pi_1} \dots$

MC: a broad class of computational algorithms that rely on repeated sampling to obtain numerical results

Monte Carlo Action Value Estimation

To estimate $V_{\pi}(s)$ for a given π

Input: π to be evaluated

Initialize:

$V(s) \in \mathbb{R}$ arbitrarily

Loop

Generate an episode: following π :
 $S_0, A_0, R_1, S_1, \dots, R_T$

$G \leftarrow 0$

Loop

$G \leftarrow \gamma G + R_{t+1}$

$V(S_t) \leftarrow \text{average}(G)$

} Only State

Action Values

- We must estimate value of each action

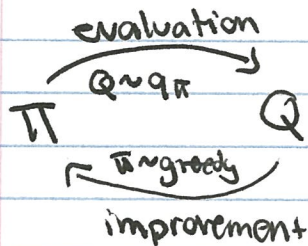
Goal: Estimate q_{π}

- we visit (s, a) pairs probabilistically

- we want to make sure every (s, a) is visited

Monte Carlo Control

How can we use MC to approximate π^*



$$\pi_0 \xrightarrow{E} q_{\pi_0} \xrightarrow{I} \pi_1 \xrightarrow{E} q_{\pi_1} \xrightarrow{I} \pi_2 \dots$$

MC with Exploring Starts to estimate $\pi \approx \pi^*$

Initialize $\pi(s) \in A(s)$
 $Q(s, a) \forall s \in S, a \in A(s)$

Loop

Choose $s_0 \in S, A_0 \in A(s_0)$ randomly

Generate Episode from $s_0 A_0 \dots R_T$

Get G

$Q(s_+, A_+) \leftarrow \text{Average}(s_+, A_+)$

$\pi(s_+) \leftarrow \max_a Q(s_+, a)$

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964

1-1-1964