Class 14

Shikhar Saxena

March 14, 2023

Contents

Recap

We will go over RL algorithms (not necessarily dwelling on *just* proofs from hereforth).

Their can be the following cases:

- State Space
  - ⋆ Discrete
  - ⋆ Continuous
- Action Space
  - ⋆ Discrete
  - ⋆ Continuous

Transition from one state to another (might be deterministic or probablistic).

For all these cases we can apply ***model-based methods***. Example MC methods, TD learning, SARSA, Q-learning. SARSA and Q-learning are special cases of Temporal Difference (TD) Learning. These are parts of Stochastic Approximation techniques.

Continuous State and Action Space

- Model based Methods
  - ⋆ Linear Quadratic Regulator (LQR)
  - ⋆ Controllability and Stability
  - ⋆ State Feedback Control
  - ⋆ Riccati Equation
- Model Free Methods

- ⋆ Function Approximation
- ⋆ Actor-Critic Methods
- ⋆ Integral RL
  - Analogous to TD Learning (but for Continuous Space)
- ⋆ Policy Gradient Methods

Definition 1 (Performance Index (PI))**.** *PI is defined as the summation of reward (a number that quantifies instantaneous change in the PI) for the entire time.*

Stochastic Dynamic Programming (SDP)

2 states and 2 actions.

$$P_{a_1} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \quad R_{a_1} = \begin{bmatrix} 11 & -4 \\ -14 & 6 \end{bmatrix}$$

$$P_{a_2} = \begin{bmatrix} 0.1 & 0.9 \\ 0.8 & 0.2 \end{bmatrix} \quad R_{a_2} = \begin{bmatrix} 45 & 80 \\ 1 & -23 \end{bmatrix}$$

$$Q(s_i, a_k) = \sum_{j=1}^{n} P_{ij}(a_k) \left(r_i(a_k, s_i, s_j) + V(s_j) \right)$$

Monte Carlo Control

- Good for episodic tasks
- Try to get expected value of $Q(s_i, a_k)$
- Start with a policy and try to explore a few steps
- At the end of each episode update policy π

Monte Carlo Control Explained.