Summary of Temporal-Difference Methods

Friday, December 4, 2020   10:41 PM

Temporal - Difference Methods

- Whereas Monte Carlo (MC) prediction methods must wait until the end of an episode to update value-function estimate (Q-table).

- Temporial - Difference (T-D) methods update the value function after every step.

TD Control

- Sarsa (0): On-policy TD control method.
  - updates at every time step to using ϵ-greedy principle

- Sarsamax (or Q-learning): - off-policy TD control Method.
  - updates at every time-step using greedy principle.

- Expected Sarsa :- on-policy TD control Method

  — updates at every time-step while considering the weighted sum of all potential actions

Analyzing Performance

- On-policy TD control Methods have better online performances (Sarsa & Expected Sarsa)

- Expected Sarsa generally achieves better performance than Sarsa.