Autonomous Networking

Gaia Maselli
Dept. of Computer Science

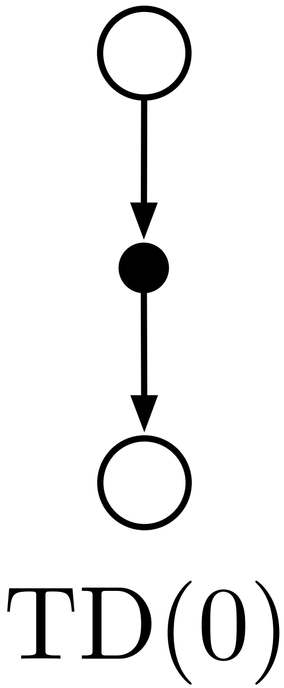
Today's plan

- Exercises

Exercise 1

- Draw and explain the backup diagram for Temporal difference learning TD(0)

Solution exercise 1



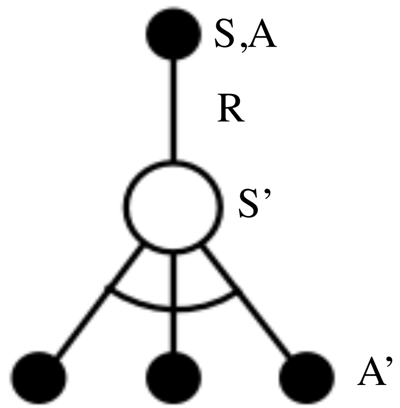
- The value estimate for the state node at the top of the backup diagram is updated on the basis of the **one sample transition** from it to the immediately following state

$$V(S_t) \leftarrow V(S_t) + \alpha\Big[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)\Big]$$

Exercise 2

- Draw and explain the backup diagram for Q-learning

Solution exercise 2



$$Q(S,A) \leftarrow Q(S,A) + \alpha\left(R + \gamma \max_{a'} Q(S',a') - Q(S,A)\right)$$

Iterative Policy Evaluation

- Iterative Policy Evaluation is a method to estimate the value function $V_{\pi}(s)$ for a given policy π.
- Goal: Compute the expected cumulative reward starting from state s while following π
- Bellman Expectation Equation:

$$V_{\pi}(s)=\sum_{a}\pi(s,a)\sum_{s'}P(s'|s,a)[R(s,a,s')+\gamma V\pi(s')]$$

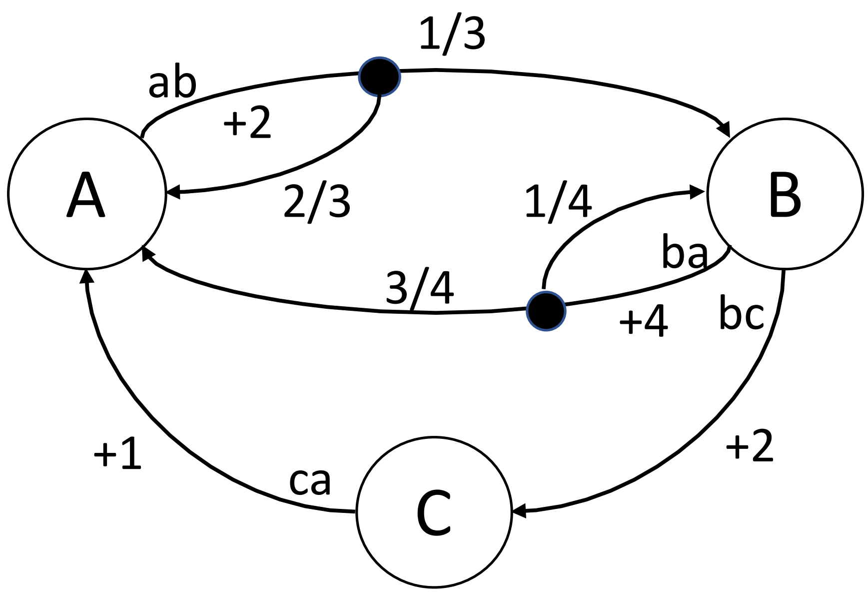
- Approach:
  - Start with an initial guess for V(s)
  - Refine the values iteratively using the Bellman equation until convergence.

Steps

- Initialization: Start with arbitrary values for $V(s)$ (e.g., $V(s)=0$).
- Iterative Update:
- $V^{\pi}(s)= \sum_a \pi(s,a) \sum_{s'} P(s'|s,a)[R(s,a,s')+\gamma V\pi(s')]$
- Repeat Until Convergence: Stop when the change between iterations is below a threshold δ.
- Result: At convergence, $V(s)\approx V^{\pi}(s)$, the true value function for the policy.
- **Synchronous Updates**:
  - Update the value for all states simultaneously using the values from the previous iteration. This is commonly used and straightforward to implement.
- **Asynchronous Updates:**
  - Update the value for states one at a time, in a specific order (e.g., topologically sorted, randomly, or sequentially).
  - Asynchronous methods can sometimes converge faster since updates can immediately use the most recent value estimates of other state

Exercise 3

- Consider the MDP with discount factor γ=0.5, with uniform random policy $\pi_1(s,a)$ that takes all actions from state s with equal probability. Starting with an initial value function of $V_1(A) = V_1(B) = V_1(C) = 1$ apply one iteration of iterative policy evaluation to compute a new value function $V_2(A)$



Solution exercise 3

$$v_\pi(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v_\pi(s') \right)$$

- $V_2(A) = 2 + 0.5*(1/3 * V_1(B) + 2/3 * V_1(A)) = 2.5$
- $V_2(B) = ?$
- $V_2(B) = 0.5*(2+0.5* V_1(C)) + 0.5*(4+0.5*(1/4*V_1(B) + 3/4 * V_1(A)))$

  $= 0.5*2.5 + 0.5 * 4.5 = 3.5$