An Introduction to Reinforcement Learning

Dongyan Lin

Imbizo 2023

1 Basic Concepts

Answer in your own words:

- What is reinforcement learning? How is it different from supervised learning and unsupervised learning?

- Draw a diagram demonstrating the interaction between the environment and the agent. Specify on your diagram: what does the agent receive from the environment? what does the environment receive from the agent?

2 Markov Decision Process

- What does it mean for a state S_t to satisfy the Markov property? Write down both the mathematical definition and the verbal meaning.

- Below are the variables that define a Markov Decision Process, What does each of them denote?
  - $\mathcal{S}$
  - $\mathcal{A}$
  - $\mathcal{P}$
  - $\mathcal{R}$
  - γ
- What is the mathematical definition of the return?
  - $G_t =$
- What is the mathematical definition of the value functions?
  - state-value $v(s) =$
  - state-action-value $q_\pi(s, a) =$
- What is the mathematical definition of the policy?
  - $\pi[a|s] =$
- (Optional) Note: not all RL problems are Markovian (in fact, the Markov property is a very hard constraint to satisfy). Can you think of an example that involves learning through trial-and-error, but does not satisfy the Markov property?

3 The Environment

- The reward function $R(s)$ is defined as the reward R the agent receives upon leaving the state s. It is usually represented as a 1-dimensional vector. What is the shape (i.e., length of each dimension) of the vector $R(s)$?

- The transition function $P(s, s', a)$ is defined as the probability of moving from state s to s' when the agent takes action a. It is usually represented by a 3-dimensional tensor $P(s, s', a)$. What is the shape of the tensor $P(s, s', a)$?

- (Optional) Sometimes the reward R that the agent receives upon leaving the state s also depends on the action it takes, a. In this case, the reward function becomes a 2-dimensional matrix, $R(s, a)$. What is the shape of the matrix $R(s, a)$?

4 The Agent

- Explain in your own words: what are the differences between model-free and model-based reinforcement learning? Hint: it might help to first answer the question: what is a "model"?

- Greedy and ϵ-greedy policy
  - What does it mean to follow a greedy policy?
    - $a_t =$

    What is the mathematical definition of a greedy policy?
    - $\pi(s) =$
  - What does it mean to follow a ϵ-greedy policy?
    - $a_t =$

    What is the mathematical definition of a ϵ-greedy policy?
    - $\pi[a|s] =$
- Describe the 3 ways to compute value of the current state, $v_\pi(s)$.

- (Optional) Derive the incremental update rule for learning value, $V_t = V_{t-1} + \alpha(G(\tau_N) - V_{t-1})$, from the equation for calculating value by sampling trajectories, $V_t = \frac{1}{N}\sum_{i=1}^{N} G(\tau_i)$.

- Describe the learning rule and draw the one-step look-ahead diagram for Q-learning and SARSA.