Reinforcement Learning

Optimality

Dr. Alireza Aghamohammadi

Action-Value Function

- The action-value function, or Q-function, measures the expected return when:
  - Starting in state s,
  - Taking action a,
  - And following policy π thereafter.
- Mathematically, it is defined as:

$$\begin{aligned} q_\pi(s,a) &= \mathbb{E}_\pi \left[G_t \mid S_t = s, A_t = a \right] \\ &= \mathbb{E}_\pi \left[R_t + \gamma G_{t+1} \mid S_t = s, A_t = a \right] \\ &= \sum_{s',r} P_r(s', r \mid s, a) \left[r + \gamma V_\pi(s') \right], \quad \forall s \in S,\ \forall a \in A(s) \end{aligned}$$

- This recursive relationship is known as the Bellman equation for action values.

Action-Advantage Function

- The action-advantage function, or simply the advantage function, measures how much better it is to take action a in state s compared to the average action under policy π:

$$a_\pi(s, a) = q_\pi(s, a) - V_\pi(s)$$

- It quantifies the relative benefit of action a over others, as determined by the policy π.

Optimality

- Optimality in reinforcement learning refers to achieving the best possible policies, state-value functions, action-value functions, and advantage functions.
- The optimal state-value function, $V^\star(s)$, gives the maximum expected return achievable from state s under any policy:

$$V^\star(s) = \max_\pi V_\pi(s)$$

- Similarly, the optimal action-value function, $q^\star(s, a)$, provides the maximum expected return for taking action a in state s:

$$q^\star(s, a) = \max_\pi q_\pi(s, a)$$

- Knowing $q^\star(s, a)$ allows us to derive the optimal policy:

$$\pi^\star[a \mid s] = \operatorname*{argmax}_a q^\star(s, a)$$