Reinforcement Learning Training 2025

Deep Q-Learning (DQN)

Update Equations

- Q-learning (tabular)

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \cdot \left[R_{t+1} + \gamma \cdot \max_{A_{t+1}} Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t) \right]$$

- Deep Q-learning

$$w_{t+1} = w_t + \alpha \,.\, \frac{1}{N} \sum_{\mathrm{i}=1}^{N} \left[r_{\mathrm{i}} + \gamma \max_{a_i'} \tilde{q}(s_i', a_i'; w_t^-) - \hat{q}(s_i, a_i; w_t) \right] . \nabla_{w_t} \hat{q}(s_i, a_i; w_t)$$

DQN

- No theoretical guarantee of convergence under non-linear function.

Components

- Policies
  - *Behavior policy*: ϵ-greedy policy to explore and generate samples.
  - *Target policy*: deterministic greedy policy (no exploration)
- Networks
  - Target network:
  - Primary network:
- Replay buffer