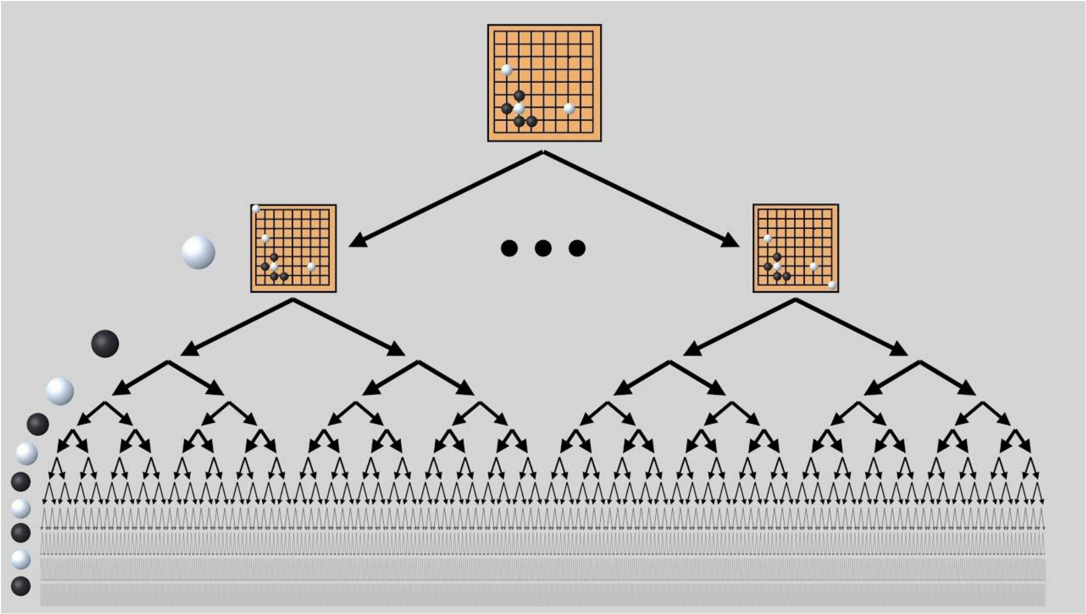
Reinforcement Learning: Function Approximation

AI/ML Teaching

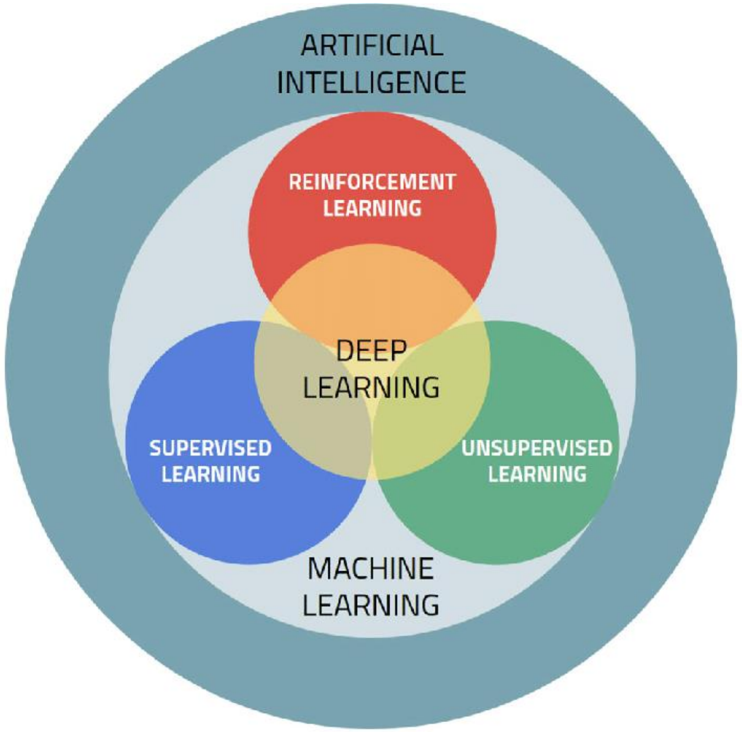
Motivation

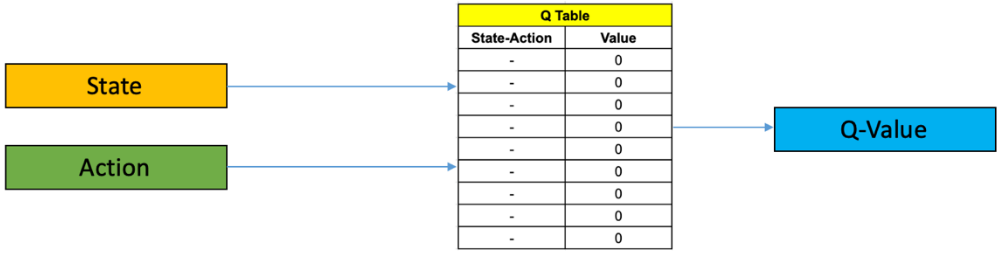
- There are too many states and/or actions to store in memory
- It is too slow to learn the value of each state individually



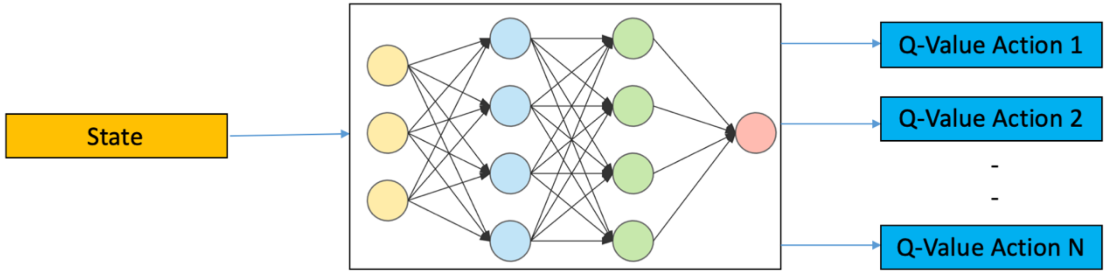
Deep learning & Reinforcement Learning

- Estimate value function with function approximation
  - $\hat{v}(s, \mathbf{w}) \approx v_\pi(s)$ / $\hat{q}(s, a, \mathbf{w}) \approx q_\pi(s, a)$
  - Generalize from seen states to unseen states
  - Neural network as a universal function approximator



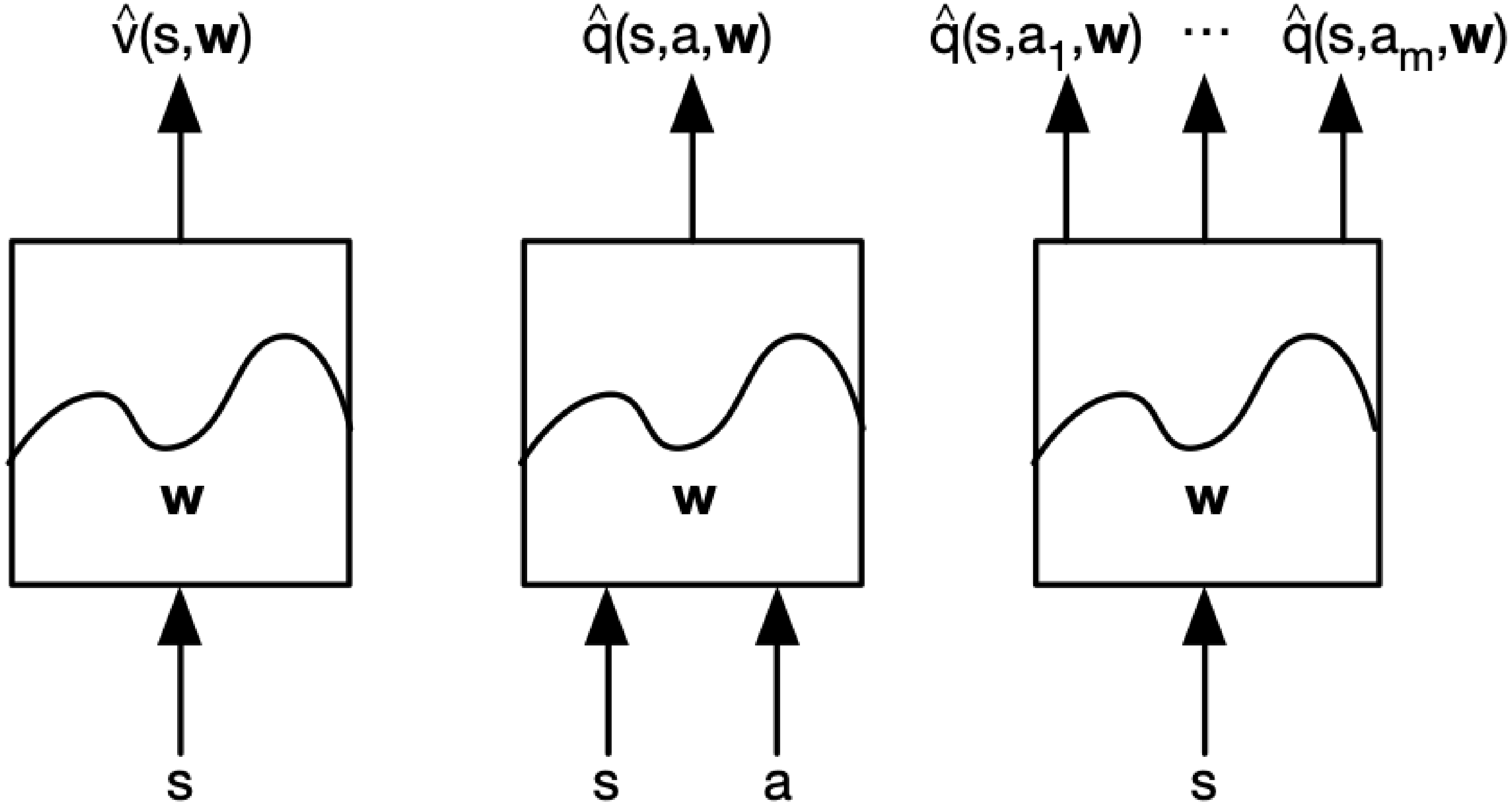


Q Learning



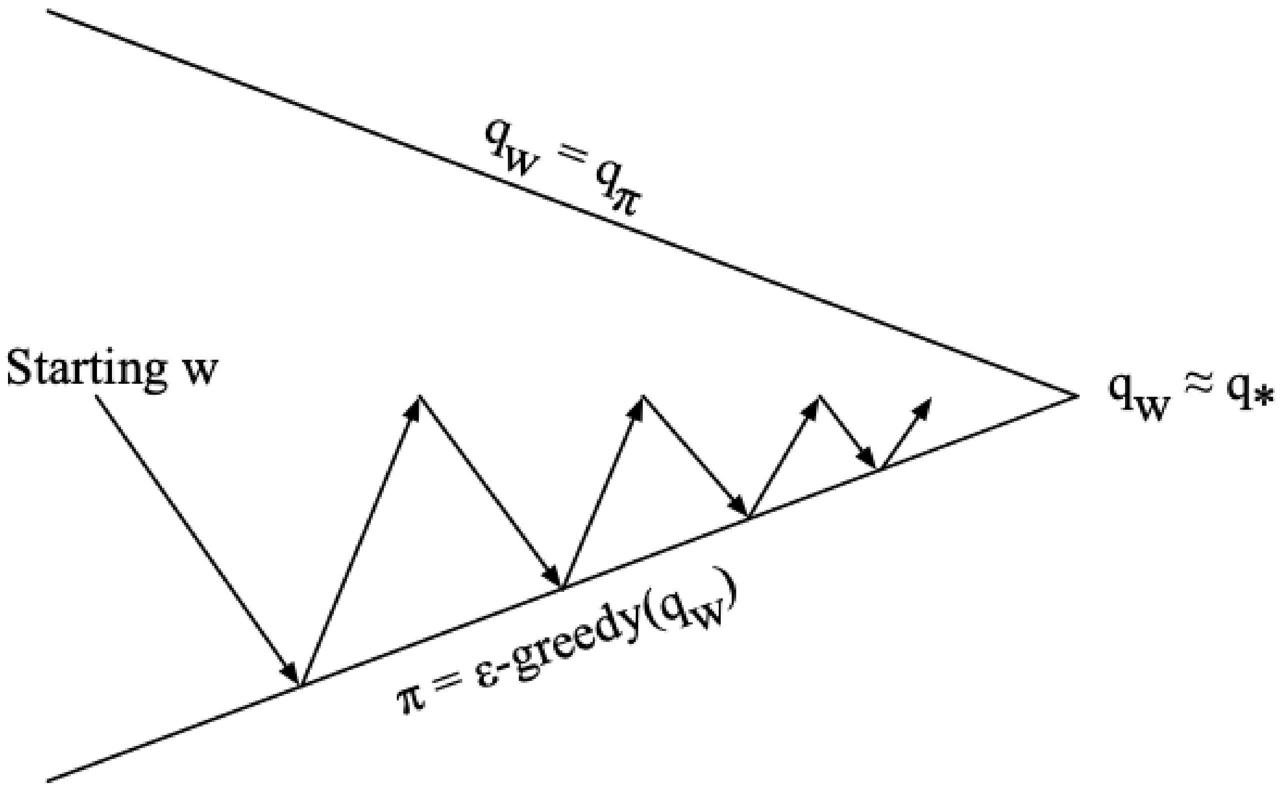
Deep Q Learning

Types of Value Function Approximation



Policy Iteration with approximate evaluation

- Policy evaluation: approximate policy evaluation
- Policy improvement: ϵ-greedy



Batch methods to RL (like supervised learning)

- Given value function approximation $\hat{v}(s, \mathbf{w}) \approx v_\pi(s)$
- Experience $\mathcal{D} = \{\langle s_1, v_1^\pi \rangle, \langle s_2, v_2^\pi \rangle, \dots, \langle s_T, v_T^\pi \rangle\}$
- Repeat:
  - Sample state, value from experience
    $$\langle s_1, v_1^\pi \rangle \sim \mathcal{D}$$
  - Apply stochastic gradient descent update
    $$\Delta \mathbf{w} = \alpha\big(v^\pi - \hat{v}(s, \mathbf{w})\big)\nabla_{\mathbf{w}} \hat{v}(s, \mathbf{w})$$
- $\mathbf{w}^\pi = \arg\min LS(\mathbf{w})$

- DQN uses **experience replay** and **fixed Q-targets**

Reference

- David Silver, COMPM050/COMPGI13 Lecture Notes
- Richard S. Sutton and Andrew G. Barto, “Reinforcement Learning: An Introduction,” 2nd Ed.