

Approaches for solving RL problems

- The Policy π : the agent's brain

- The policy π is the brain of our agent. It's the function that tells us what action to take given the state we are in.

$$\text{State} \rightarrow \pi(\text{State}) \rightarrow \text{Action}$$

- There are 2 ways to train our agent:

- Directly, by teaching the agent to learn which action to take given the current state: Policy-based methods

- Indirectly, teach the agent to learn which state is more valuable.

- Policy-Based Methods

- In Policy-Based Methods, we learn a policy function directly.

- Types of policies:

- Deterministic: a policy at a given state will always return the same action.

$$a = \pi(s)$$

- Stochastic: outputs a probability distribution over actions.

$$s_0 \rightarrow \pi(A | s_0) \rightarrow [\text{Left: } 0.1, \text{Right: } 0.9]$$