Reinforcement Learning
Trainer: Prashant N.

Agent Environment Interaction

$*$ POLICY_FUNCTION $\}\rightarrow$ This function helps the agent to decide appropriate action based on the provided state.

$S_i \uparrow \quad \downarrow A_i$

reward
R_0

→ AGENT —

State / Observation
S_0
S_1

env.reset()

Perform Action
A_0
env.step (action)

isTerminated? ———→ false: continue

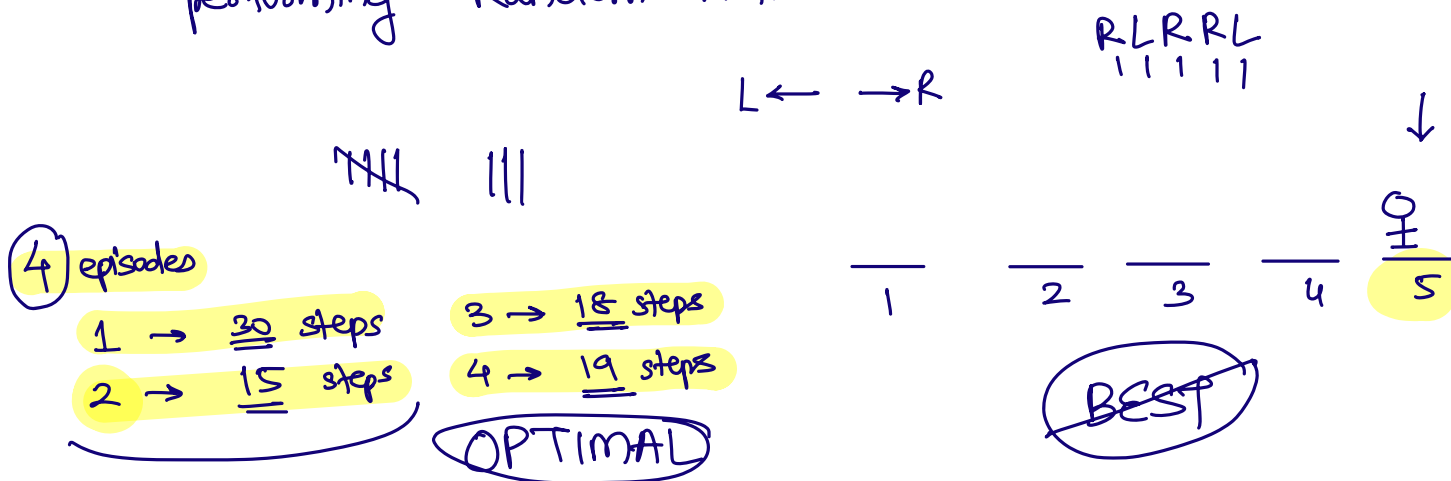true: Terminate.

ENVIRONMENT ←——
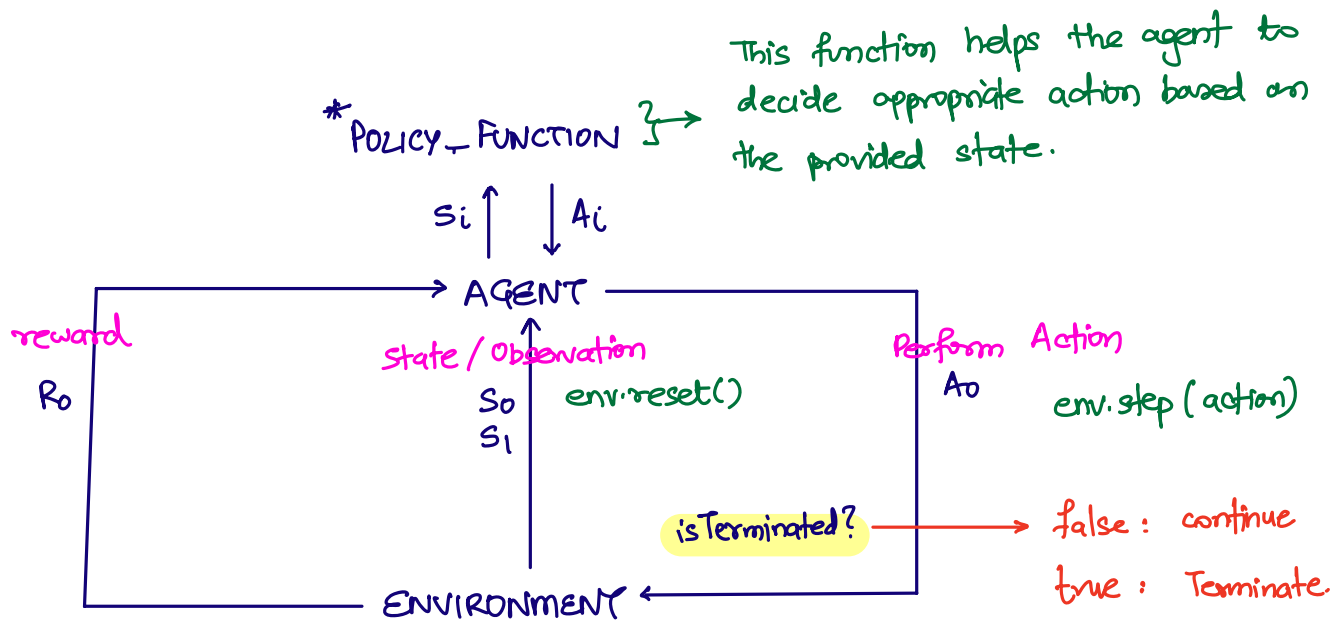
Policy function

Two core concepts of RL. w.r.t how agent decides which action to take at each step in learning.

① Exploration.

FINDING ALL POSSIBILITIES to reach the goal by performing Random Action.

$L \leftarrow \quad \rightarrow R$

R L R R L

$\overline{1} \quad \overline{2} \quad \overline{3} \quad \overline{4} \quad \underline{\overset{\downarrow}{\underset{5}{\text{Ω}}}}$

BEST

④ episodes
1 → 30 steps   3 → 18 steps
2 → 15 steps   4 → 19 steps

OPTIMAL

② Exploitation

Applying ==KNOWN SOLUTION== to reach the Goal.

-or-

Using a ==Guidance Mechanism== to reach the Goal.

As a programmer following things will happen:

① You know the optimal solution and hard code the same
$$[R, R, R, R]$$

$$\overline{} \quad \overline{} \quad \overline{} \quad \overline{} \quad \overline{}^{\;♀}$$
$$1 \qquad 2 \qquad 3 \qquad 4 \qquad 5$$

② You load an ==optimized== model to reach the solution