

Hado Van Hasselt, Deep Mind

MC: Monte Carlo

TD ~~temporal~~

总结: 强化学习

Supervised Learning: Function Approximation

(Approximate Function Induction)

(data $\xleftrightarrow{\text{mapping}}$ labels)

Unsupervised Learning: Compact Description

(Finding structure hidden
(in unlabeled data))

Reinforcement Learning: Learning to maximize expected
long-term reward

(data \rightarrow action \rightarrow long-term reward)

FOUR subelements:

policy: learning agent's way of behaving at a given time.

policy
State $\xrightarrow{\quad}$ action
mapping

"Stimulus - response"

(function / lookup table / search process)

Reward signal: goal of RL problem

"experiences of pleasure or pain"

(function f | State of the environment
action)

Value function: What is good in the long run

(total amount of reward an agent can expect to accumulate over the future,

~~starting from~~ that state)

Model of the Environment:

allow inferences to be made about how the environment will behave