

* Fully vs Partially Observation

(전체 정보) ↓

(ex: camera vision) ↓

* Model based

(next reward & state transition)

vs Free

(Agent modeling)

* Stochastic vs Deterministic policy

($a = \pi(s)$)

* Value Based vs Policy Based vs Actor Critic

Using value

policy

Both

* Reinforcement Learning vs Planning

(Unknown env)

(known env)

(next reward &

state transition)

* Exploration vs Exploitation

(정보 추가)

(정보 활용)

* Prediction vs Control

(Given policy)

(Find best policy)

(Train value)

(policy가 주어지기 때문에 해당 policy에 대해

value function을 구하는 것이지, 모든 policy에 대해

Optimal value function을 구하는 건 아님)

* Action을 할 확률과.

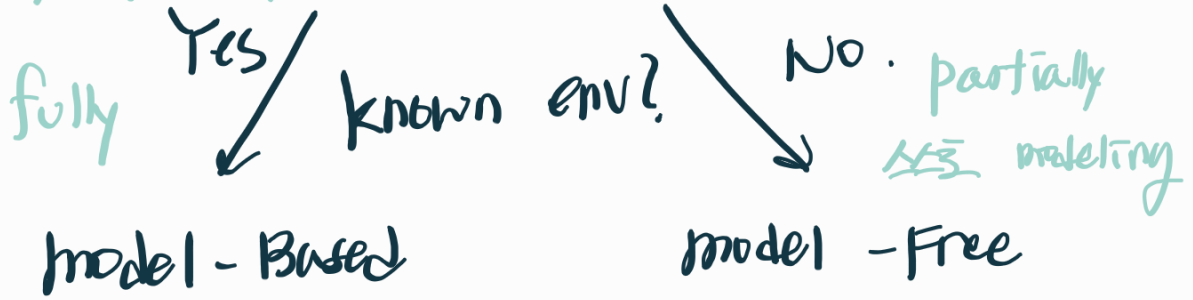
* $\pi(s) \rightarrow A$

action을 했을 때 해당 state로

갈 확률은 서로 다른 개념이다.

* Problem Diagram.

MDP < S, P, R, γ , A > **problem**



planning

(dynamic programming) ⇓



Reinforcement Learning

Using value-Based / policy-Based / actor-critic



prediction or Control
(estimate) (optimize)

+ Fully or partially observation?
stochastic or deterministic policy?
(exploration & exploitation)

