Page No.
Date : / /20

#20days of #100days of learning.

Reienforcement learning ⇒ "If you ask the wrong question, you will never get the right answer".

* Supervised
Given x, predict y

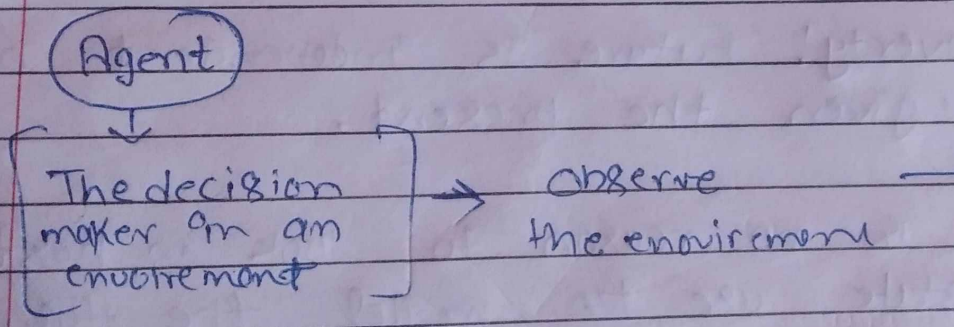
Unsupervised
Given x, Simplify x

* Neither supervised nor unsupervised learning will work in an unknown environment.

Definition ⇒

Train decision makers to take actions to maximize rewards in uncertain environment.

e.g. How long self drive car go without accident



Reinforcement Learning	Supervised / Unsupervised Learning
Objective: choice "best" actions	objective: Predict, Classify or Simplify
Environment is uncertain	Environment is known
Training involves exploring the environment	Training involved finding Pattern in data.
Training process involved determining the best policy	Training process involves fitting the "best models"
Explicit dependency of rewards on previous action	Individual point are independent of each other

* Environment as a Markov Decision Process (MDP)

Markov Property: Future is independent of the past, given the present.

⇒ There is no reason to look in past. Present state use to model the future.

Steps involve in MDP ⇒ At each time step

1. ~~environment in some state~~ ⇒ S_t
   ↳ Decision maker can choose an action ⇒ a
2. moves environment to new state ⇒ S_{t+1}
   ↳ Decision maker receives reward ⇒ $R_a(S_t, S_{t+1})$
3. S_{t+1} depends only a and S_t.

Page No.
Date : / /20

Markov Property ⇒

Future (S+1) is independent of the past (S+1, S+2 ...), given the present (a, St)

* Policy Search Algorithm ⇒

① Brute force methods → Evaluate every possible policy over every possible state
② Policy Gradient methods → Explore state space to find best policy
③ Value function methods → • Explicitly model environment as Markov Decision process.
• Popular and robust.

Several Implementation ⇒

- Q-learning
- SARSA
- Monte Carlo