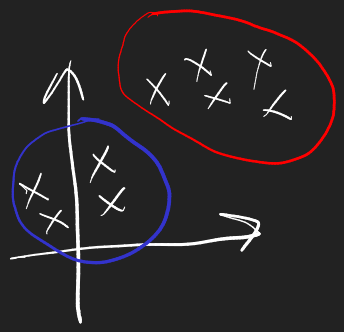
1) What is RL? ✓
2) Logistics ✓
3) Types of Machine Learning

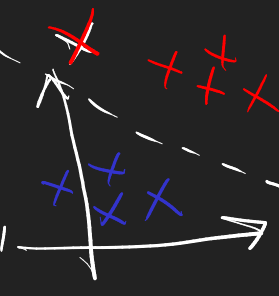
1) Unsupervised learning
   → goal: summarize
   → dataset: $\{x_i\}_{i=1}^{N}$
   → evaluate: qualitative "descriptive"



ex- PCA

K-means clustering

2) Supervised learning
   → goal: prediction
   → dataset: $\{(x_i, y_i)\}_{i=1}^{N}$ (features, labels)



"predictive"

→ evaluation: accuracy $\hat{y}$ vs. y

3) Reinforcement Learning

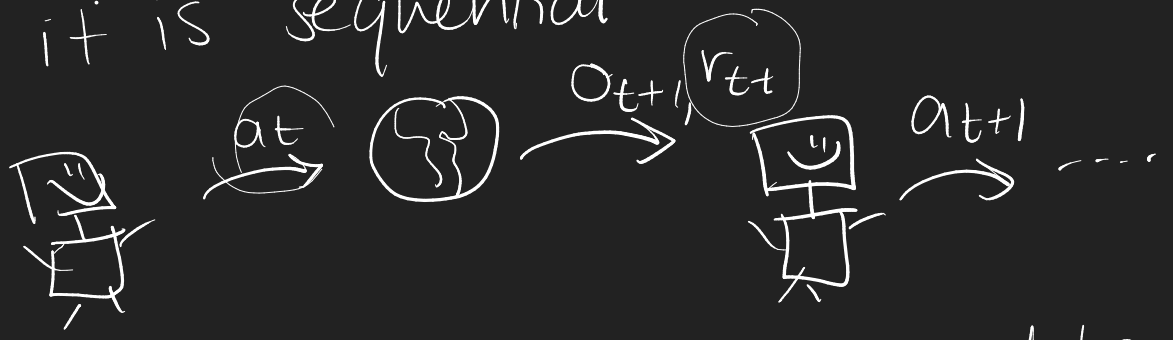
→ goal: action / decision

→ data set: $\{(o_t, a_t, r_t)\}_{t=0}^{N}$ ← sequential

observation, action, reward

→ evaluation: cumulative reward

Unlike supervised / unsupervised learning, data is not drawn iid, instead it is sequential

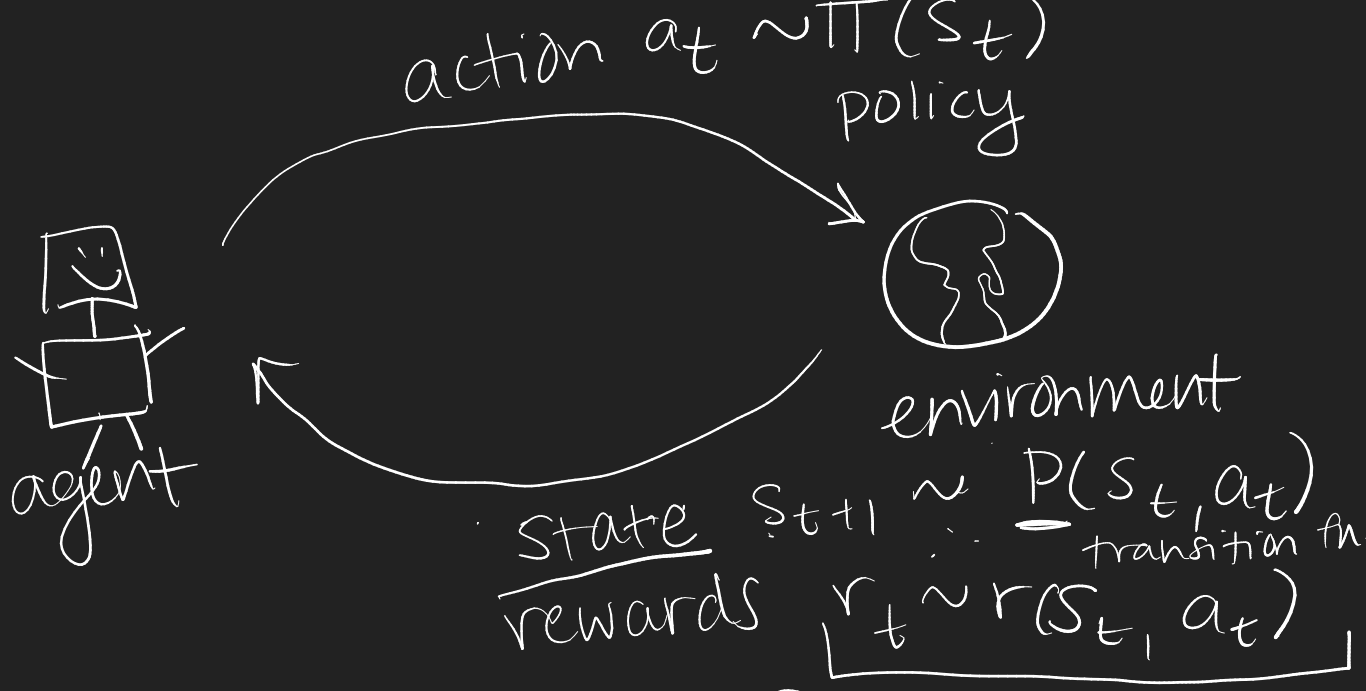


1) may start with no data

2) actions have consequences affect future data

3) Solving a task requires a long sequence of actions

4) Markov Decision Processes (MDP)



Structure in MPP

environment has state s_t

Markovian assumption:

$$P\{s_{t+1} = s \mid s_t, s_{t-1}, \cdots, s_0, a_t, a_{t-1}, \cdots, a_0\}$$

$$= P\{s_{t+1} = s \mid s_t, a_t\}$$

In this class, we usually assume

$$o_t = s_t$$

Example manipulation

state s: configuration of hand, object pose (acceleration/velocity)

action a: motor position commands (torques)

transition: $s' \sim P(s,a)$ physics (gravity, contact force, friction)

policy $\pi(s)$: map from config. to motor commands

Reward $r(s,a)$: · negative distance to goal pose
· negative motor command magnitude

Q: if there are S ($\in \mathbb{Z}^+$) states and A ($\in \mathbb{Z}^+$) actions, how many deterministic policies are there?

π: states $\rightarrow$ actions

~~$S \times A$~~

~~A^S~~

A^S