Reinforcement learning cont

→ Actor critic (Both)

★ Ex's of RL: Games (Alpha Go, Chess), Robotics,

Finance (Stock predictions), Health care, self driving cars (tesla)
↳ ex of Policy based (Learn stratigy) ↳ Value based (learn Best moves)

★ Algos: Q-learning, DQN, SARSA, PPO, A3C, REINFORCE, etc

★ challanges: sample inefficiency → Needs lots of trail and error

Exploration is hard → can get stuck on bad habits

sparse rewards → sometimes the agent rarely gets Feedback

Safty → in real world (cars) bad action can be dangourous

★ analogy: imagine training a dog

- state: Dog sees ball • action: Dog fetches or ignores
- Reward: Gets reward for feching ▫ over time dog learns Fetch = treat ⇒ reinforced behaviour

★ Training (a training loop is trail and error learning):

1) initialize agent and enviorment
- agent starts with random policy (no clue what to do)
- Env is set (eg a game, sim robot)

2) interactive loop
- agent observes state (s) • agent chooses action (a) random at first later learned from policy
- Env responds with: • Reward (r) → (numaric Feedback) • Next state (s') → new situation

3) learning step (update step)
- Agent updates either: Value function (QL/DQL) or Policy parameter (PG) or Both (actor critic)
- goal = maximize expected cumulative reward over time

4) Repeat many episodes: the agent plays episodes (Full set of tasks till task ends)
- with each episode policy improves • Try → Feedback → Adjust → Repeate