(10) Reinforcement Learning II

★ How to Explore?

⇒ Several Schemes for forcing exploration

Simplest: random action (ε-greedy)

- Every time step, flip a coin
- With (small) probability ε, act randomly
- With (large) probability 1-ε, act on current policy

⇒ Problem with random actions?

- You do eventually explore the space, but keep thrashing around once learning is done.
- One Solution: lower ε over time
- Another Solution: Exploration functions

★ Exploration Function

- Random actions: Explore a fixed amount
- Better idea: Explore areas whose badness is not (yet) established, eventually stop exploring

⇒ Takes a value estimate u and a visit count n & returns an optimistic utility:

$$f(u,n) = u + k/n$$

Modified Q-update: $Q(s,a) \xleftarrow{\alpha} R(s,a,s') + \gamma \max_{a'} f\left(Q(s',a'), N(s',a')\right)$

★ Regret

⇒ Measure of your (total mistake cost)

Difference between your rewards

& optimal rewards

$$\text{regret} = u(\text{optimal action}) - u(\text{action taken})$$