■ MDP Cheat Sheet (Lecture 1)

MDP Definition

An MDP is a 4-tuple (S, A, R, P):
• S: set of states
• A: set of actions
• R(s,a) or R(s,a,s'): reward model
• P(s,a,s'): transition probabilities
Markov property: Future depends only on (s,a), not the full past.

Policies

• Deterministic: $\pi(s) \in A$
• Randomized: $\pi(s) \in \Delta(A)$
• Stationary: same rule forever
• Non-stationary: rule changes over time

Objective Functions

Finite Horizon	V^π(s1) = E[Σ_{t=1}^H γ^(t-1) r_t]
Infinite Horizon (Discounted)	V^π(s1) = E[Σ_{t=1}^∞ γ^(t-1) r_t]
Infinite Horizon (Average)	ρ^π(s1) = lim (1/T) Σ_{t=1}^T E[r_t]

Value Functions

• V^π(s): expected return starting from s under π
• Q^π(s,a): expected return starting from (s,a) then following π

Bellman Equations

Policy Evaluation:
V^π(s) = E[R(s,a) + γ Σ_{s'} P(s,a,s') V^π(s')]
Optimal Value:
V*(s) = max_a [R(s,a) + γ Σ_{s'} P(s,a,s') V*(s')]
Optimal Q-value:
Q*(s,a) = R(s,a) + γ Σ_{s'} P(s,a,s') max_{a'} Q*(s',a')

Algorithms to Solve MDPs

• Value Iteration: iteratively apply Bellman optimality updates
• Policy Iteration: alternate evaluation and improvement
• Q-value Iteration: update Q directly
• Linear Programming: solve Bellman fixed point as LP