Fraunhofer IIS

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Introduction to Model-free RL

Christopher Mutschler

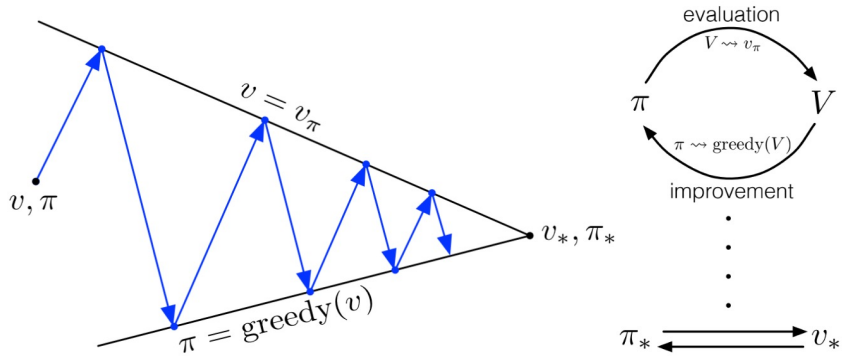
Recap: RL, MDPs, DP

- The general RL setting
  - Agent-Environment-Interface: actions, states, and rewards
  - Agent interacts with the environment over a sequence of discrete time steps (episodic or continual)
  - Policy as a stochastic rule to select actions

- MDPs as tools to describe RL problems
  - Main ingredients: states, actions, state transition probabilities, return, and discount
  - Value functions that describe the expected return following a particular policy
  - Bellman equation as expression of the relationship between the value of a state and the value of its successor states

Recap: RL, MDPs, DP

- Dynamic Programming (DP) methods to find optimal controllers
  - DP methods are guaranteed to find optimal solutions for Q and V in polynomial time (in number of states and actions) and are exponentially faster than direct search
  - Policy Iteration computes the value function under a given policy to improve the policy while value iteration directly works on the states
    - Perform sweeps through the state set
    - Implement the Bellman equation update
    - Use bootstrapping
  - Require complete and accurate model of the environment
  - Have limited applicability in practice…
    - …as they need to know the dynamics of the environment!



Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Monte Carlo and TD Methods

- So far: We know our MDP model $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$.
  - Planning by using dynamic programming
  - Solve a known MDP
- What if we don't know the model, i.e., $\mathcal{P}$ or $\mathcal{R}$ or both?

- We distinguish between 2 problems for unknown MDPs:

  - Model-free Prediction: Evaluate the future, given the policy π.
    (estimate the value function)

  - Model-free Control: Optimize the future by finding the best policy π.
    (optimize the value function)