

Policy and Value Iteration

Last Time

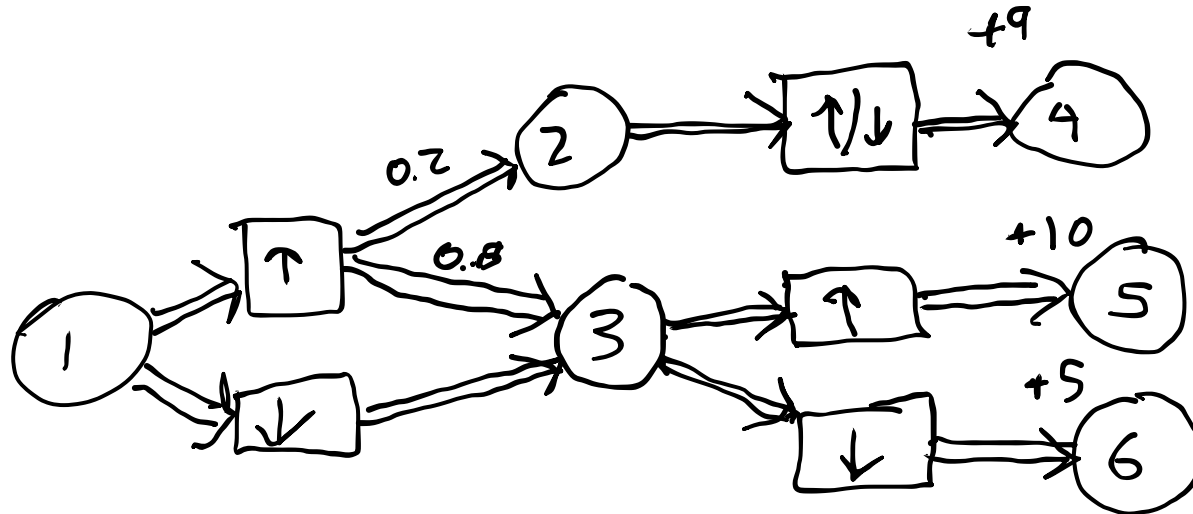
- How is a **Markov decision process** defined?
- What is a **policy**?
- How do we **evaluate** policies?

Guiding Questions

- How do we reason about the **future consequences** of actions in an MDP?
- What are the basic **algorithms for solving MDPs**?

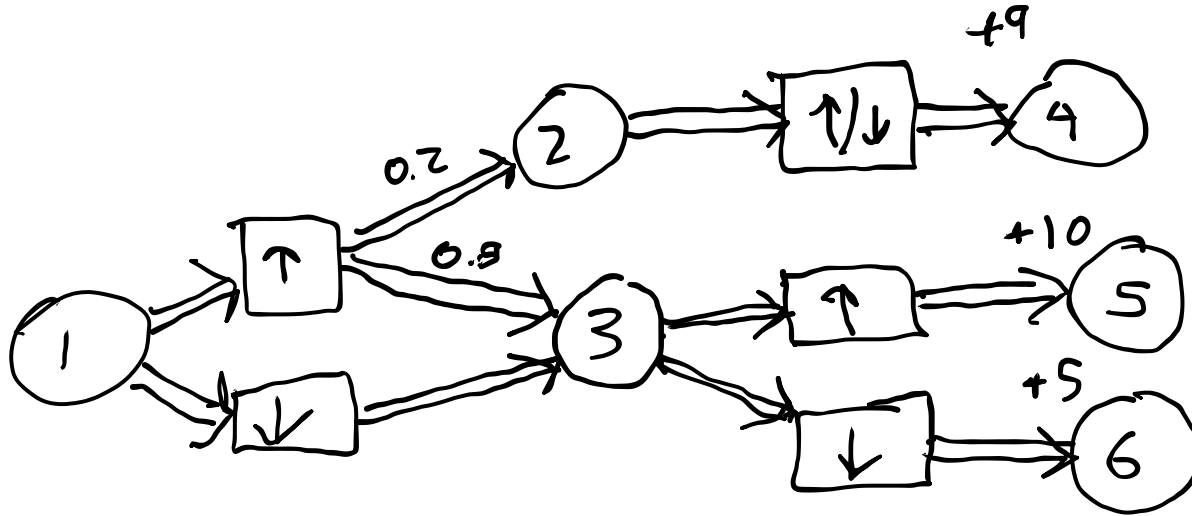
MDP Example: Up-Down Problem

For this lecture, \Rightarrow is same as \rightarrow (distinguishes from Bayes Net)



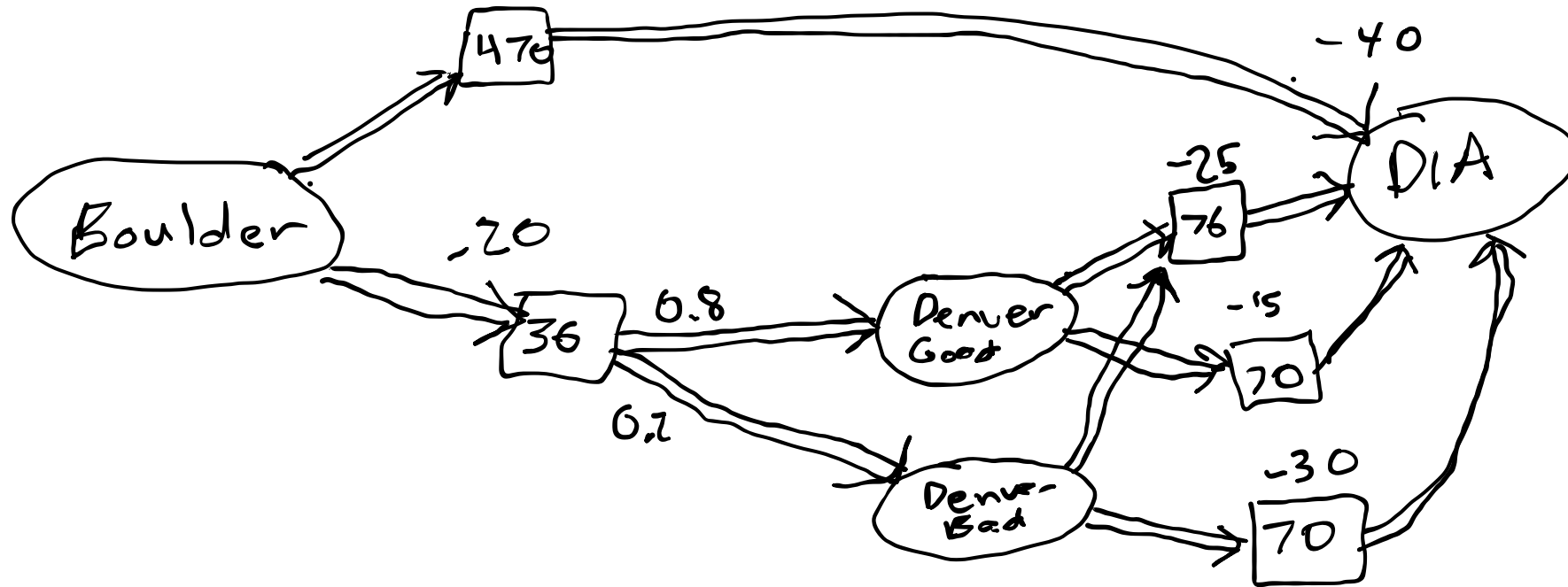
Next Year: Include example when the same state is visited again like 2023 Exam 3 Question 3c

Dynamic Programming and Value Backup



Bellman's Principle of Optimality: Every sub-policy in an optimal policy is locally optimal

Break: DIA Run



Policy Iteration

Algorithm: Policy Iteration

Given: MDP (S, A, R, T, γ, b)

1. initialize π, π' (differently)
2. while $\pi \neq \pi'$
3. $\pi \leftarrow \pi'$
4. $U^\pi \leftarrow (I - \gamma T^\pi)^{-1} R^\pi$
5. $\pi'(s) \leftarrow \operatorname{argmax}_{a \in A} (R(s, a) + \gamma \sum_{s' \in S} T(s'|s, a) U^\pi(s')) \quad \forall s \in S$
6. return π

Value Iteration

Algorithm: Value Iteration

Given: MDP (S, A, R, T, γ, b) , tolerance ϵ

1. initialize U, U' (differently)
2. while $\|U - U'\|_\infty > \epsilon$
3. $U \leftarrow U'$
4. $U'(s) \leftarrow \max_{a \in A} (R(s, a) + \gamma \sum_{s' \in S} T(s'|s, a)U(s')) \quad \forall s \in S$
5. return U'

- Returned U' will be close to U^* !
- π^* is easy to extract: $\pi^*(s) = \arg \max (R(s, a) + \gamma E[U^*(s)])$

Bellman's Equations

Guiding Questions

- How do we reason about the **future consequences** of actions in an MDP?
- What are the basic **algorithms for solving MDPs**?

"In any small change he will have to consider only these quantitative indices (or "values") in which all the relevant information is concentrated; and by adjusting the quantities one by one, he can appropriately rearrange his dispositions without having to solve the whole puzzle ab initio, or without needing at any stage to survey it at once in all its ramifications."

-- F. A. Hayek, "The use of knowledge in society", 1945