Markov Decision Processes

MDPs

- **discounted reward**: penalize future rewards by γ
  - $R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \cdots + \gamma^n R(s_n)$
- **policy**: $\pi(s) = a$ gives an action for each state
- **optimal policy**: π^*

Calculating π^*

Optimal policy is based on optimal utility:

$$\pi^*(s) = \operatorname*{argmax}_{a \in A(s)} \sum_{s'} T(s,a,s') U^{\pi^*}(s')$$

Utility of a policy is based on expected rewards:

$$U^{\pi}(s) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t) | \pi, s_0 = s]$$

Utility of a state is its reward plus the best utility of an action in that state:

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s') U(s')$$

Calculating π^*

Optimal policy is based on optimal utility:

$$\pi^*(s) = \operatorname*{argmax}_{a \in A(s)} \sum_{s'} T(s,a,s') U^{\pi^*}(s')$$

Utility of a policy is based on expected rewards:

$$U^{\pi}(s) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t) | \pi, s_0 = s]$$

Utility of a state is its reward plus the best utility of an action in that state:

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s') U(s')$$

Calculating π^*

Optimal policy is based on optimal utility:

$$\pi^*(s) = \underset{a \in A(s)}{\operatorname{argmax}} \sum_{s'} T(s, a, s') U^{\pi^*}(s')$$

Utility of a policy is based on expected rewards:

$$U^{\pi}(s) = E[\sum_{t=0}^{\infty} \gamma^t R(s_t) | \pi, s_0 = s]$$

Utility of a state is its reward plus the best utility of an action in that state:

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s, a, s') U(s')$$

Solving MDPs

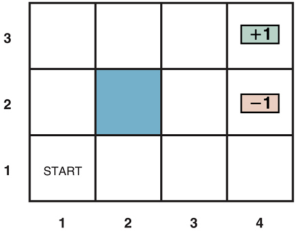
Value Iteration

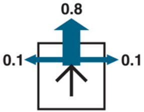
Repeated Bellman updates:

for all States s **do**
    $U(s) \leftarrow R(s)$
while unsatisfied **do**
    for each state s **do**
        $U'(s) \leftarrow R(s) + \gamma \max_a \sum_{s'} T(s,a,s') U(s')$
    $U \leftarrow U'$

Utility values are guaranteed to converge with enough updates
This equilibrium gives an optimal policy

MDP Example





Transitions to terminal states have rewards of -1 and 1, all other transition rewards are $-.04$

Probability of .8 to move in intended direction, .1 to move at a right angle

$0 < \gamma \leq 1$ – let's pick $\gamma = .5$