4.1 Policy Evaluation (Prediction)

Convert Bellman Equation into an update rule.

$$V_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \lambda V_k(s')\right]$$

Ex 4.1) $q_\pi(11, \text{down}) = -1$

$$q_\pi(7, \text{down}) = E_\pi[G_t | S_t = s, A_t = a]$$
$$= R_t + \lambda E_\pi[G_{t+1} | S_{t+1} = s']$$
$$= (-1) + (1)(-14)$$
$$= -15$$

$k = 1$

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

$$V = V(1)$$
$$= 0$$
$$V(1) = \frac{1}{4}(-1 + (1)(0)) \cdot 4$$
$$= -1$$

4.2) $V_\pi(15) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \lambda V_\pi(s')\right]$

$$\frac{1}{4}(-1 - 22 - 1 - 20 - 1 - 14 - 1 + V_\pi(15))$$

$$V_\pi(15) = -20$$

4.2 Policy Improvement

Policy improvement theorem:

$$q_\pi(s, \pi'(s)) \geq q_\pi(s, \pi(s)) \quad \forall s \in S \rightarrow \pi' \geq \pi$$

$$q_\pi(s, \pi'(s)) \geq q_\pi(s, \pi(s)) \quad \exists s \in S \rightarrow \pi' > \pi$$

state = # of parking spaces occupied

action = price of parking

reward = city preference