Reinforcement Learning Exercise

Exercises 03

1 Dynamic programming

(a)

$$\begin{aligned} q_\pi(s,a) &= \mathrm{E}_\pi[G_t \,|\, S_t = s, A_t = a] \\ &= \mathrm{E}_\pi[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s, A_t = a] \\ &= \mathrm{E}_\pi[R_{t+1} + \gamma \sum_{s',a'} q_\pi(s',a') \,|\, S_t = s, A_t = a] \\ &= \sum_{s',r'} p(s',r|s,a) \left(r + \gamma \sum_{a'} \pi(a'|s') q_\pi(s',a') \right) \end{aligned}$$

$$\begin{aligned} q_{k+1}(s,a) &= \mathrm{E}_\pi[R_{t+1} + \gamma G_{t+1} \,|\, S_t = s, A_t = a] \\ &= \sum_{s',r} p(s',r|s,a) \left(r + \gamma \sum_{a'} \pi(a'|s') q_k((s',a') \right) \end{aligned}$$