

1. 解释Q和V值函数的区别

Q值函数表示的是从状态s出发，采取行动a后继续采用策略π的收益，即 $Q(s, a)$ 是 π, s, a 的函数。

V值函数表示的是从状态s出发，采取不同行动的收益，即 $V(s)$ 是 π, s 的函数

2. 写出两个函数的bellman递推式。V值函数中不应出现动作A。

$$\begin{aligned} V_{\pi}(s) &= \sum_{s' \in \mathbf{S}} T(s, a, s') [U(s, a, s') + V_{\pi}(s')] = \\ &= \sum_{s' \in \mathbf{S}} T(s, \pi(s), s') [U(s, \pi(s), s') + V_{\pi}(s')] \\ Q_{\pi}(s, a) &= \sum_{s' \in \mathbf{S}} T(s, a, s') [U(s, a, s') + V_{\pi}(s')] \end{aligned}$$