

# v2.0 符号表

名称	符号
状态集合	$S$
动作集合	$A$
折扣因子	$\gamma \in [0, 1]$
t 时刻的状态	$s_t$
t 时刻的动作	$a_t$
t 时刻的奖励	$r_t$
状态转移概率	$P(s_{t+1}   s_t, a_t)$
奖励函数	$R(s_{t+1}, a_t, s_t)$
轨迹	$\tau = \{s_0, a_0, s_1, a_1, \dots\}$
轨迹的奖励	$R(\tau) = \sum_{t=1}^T r_t$
回报（累计折扣奖励） Reward to go	$G_t = \sum_{k=0}^{T-t-1} \gamma^k r_{t+k}$
策略	$\pi_{\theta}(a_t   s_t)$ —— 随机性策略（其中 $\theta$ 是网络参数） $\mu_{\theta}(s_t)$ —— 确定性策略（其中 $\theta$ 是网络参数）
状态价值函数	$V_{\pi}(s) = \mathbb{E}_{\pi}[G_t   s_t = s]$
动作价值函数	$Q_{\pi}(s, a) = \mathbb{E}_{\pi}[G_t   s_t = s, a_t = a]$
优势函数	$A_{\pi}(s, a) = Q_{\pi}(s, a) - V_{\pi}(s)$
轨迹出现的概率	$p_{\theta}(\tau) = p(s_0) \prod_{t=1}^T \pi_{\theta}(a_t   s_t) p(s_{t+1}   s_t, a_t)$
期望奖励	$\bar{R}_{\theta} = \sum p_{\theta}(\tau) R(\tau) = \mathbb{E}_{\tau \sim p_{\theta}(\tau)}[R(\tau)]$