常用符号

符号	中文	英文
S 或 s	状态	state
A 或 a	动作	action
R 或 r	奖励	reward
U 或 u	回报	return
γ	折扣率	discount factor
$\mathcal{S}$	状态空间	state space
$\mathcal{A}$	动作空间	action space
$\pi(a\|s)$	随机策略函数	stochastic policy function
$\boldsymbol{\mu}(s)$	确定策略函数	deterministic policy function
$p(s'\|s,a)$	状态转移函数	state-transition function
$Q_\pi(s,a)$	动作价值函数	action-value function
$Q_\star(s,a)$	最优动作价值函数	optimal action-value function
$V_\pi(s)$	状态价值函数	state-value function
$V_\star(s)$	最优状态价值函数	optimal state-value function
$D_\pi(s)$	优势函数	advantage function
$D_\star(s)$	最优优势函数	optimal advantage function
$\pi(a\|s;\boldsymbol{\theta})$	随机策略网络	stochastic policy network
$\boldsymbol{\mu}(s;\boldsymbol{\theta})$	确定策略网络	deterministic policy network
$Q(s,a;\boldsymbol{w})$	深度 Q 网络	deep Q network (DQN)
$q(s,a;\boldsymbol{w})$	价值网络	value network

目录