# 《强化学习：原理与 Python 实现》字母表

这里只列出常用字母。部分小节会局部定义字母，此时以局部定义为准。

一般规律：大写字母是随机事件或随机变量，小写字母是确定性数值或确定性变量。衬线体（如 Times New Roman 字体）是数值，非衬线体（如 Open Sans 字体）不一定是数值。粗体是向量或矩阵。花体是集合。

## 拉丁字母

| | | | |
|---|---|---|---|
| $a$，$a_\pi$ | 优势 | $q$，$q_\pi$ | 动作价值 |
| $A$ | 动作（随机事件） | $q_*$ | 最优动作价值 |
| $a$ | 动作事件 | $R$ | 奖励（随机变量） |
| $\mathcal{A}$ | 动作空间 | $r$ | 奖励值 |
| $b$ | 行为策略 | $\mathcal{R}$ | 奖励空间 |
| $B$ | 策略梯度中的基线（随机量） | $\mathbb{R}$ | 实数集 |
| $\mathcal{B}$ | 经验回放中抽取的一批经验 | $S$ | 状态（随机事件） |
| $c$ | 计数值；线性规划的目标系数 | $s$ | 状态事件 |
| $d$，$d_\infty$ | 度量 | $\mathcal{S}$ | 状态空间 |
| $d_{\mathrm{KL}}$ | KL 散度 | $T$ | 回合步数（随机变量） |
| $\mathcal{D}$ | 经验回放中的经验集 | $t$ | 时间指标 |
| $e$ | 资格迹 | $(\ )^{\mathrm{T}}$ | （矩阵的）转置 |
| E | 期望 | $U$ | 用自益得到的回报估计 |
| $G$ | 回报（随机变量） | $V$ | 状态价值估计（随机变量） |
| $g$ | 回报值 | $v$，$v_\pi$ | 状态价值 |
| $\mathbf{g}$ | 梯度向量 | $v_*$ | 最优状态价值 |
| $h$ | 动作偏好；熵 | $\mathbf{w}$ | 价值估计参数 |
| $k$ | 迭代步数 | $X$ | 一般的随机事件 |
| $\mathbb{N}$ | 自然数集 | $x$ | 一般的事件 |
| $p$ | 概率值 | $\mathcal{X}$ | 一般的事件空间 |
| Pr | 概率 | $\mathbf{z}$ | 资格迹参数 |
| $Q$ | 动作价值估计（随机变量） | | |

## 希腊字母

| | | | |
|---|---|---|---|
| $\alpha$ | 学习率 | $\pi$ | 策略 |
| $\beta$ | 资格迹算法强化强度 | $\pi_*$ | 最优策略 |
| $\gamma$ | 折扣因子 | $\boldsymbol{\theta}$ | 策略估计参数 |
| $\delta$ | 时序差分误差 | $\vartheta$ | 价值迭代终止阈值 |
| $\varepsilon$ | 探索参数 | $\rho$ | 重要性采样比率 |
| $\lambda$ | 资格迹衰减强度 | $\Psi$ | 扩展的优势估计（随机变量） |

## 其他符号

| | |
|---|---|
| $\leq$ | 普通数值比较；向量逐元素比较；策略的偏序关系 |
| $\ll$ | 绝对连续 |