RL Lecture: Notation Cheat Sheet

M. Lindauer – Leibniz University Hannover

Symbol	Meaning
S	space of states
$s \in S$	one specific state
A	space of Actions (**Warning:** overloaded notation with advantage function)
$a \in A$	one specific action
$P : S \times A \to S$	dynamics of environment
$R : S \times A \to \mathbb{R}$	reward function
$r \in \mathbb{R}$	concrete reward value
$\gamma \in [0, 1]$	discount factor
T	maximal time horizon (**Warning:** overloaded notation with terminal states)
t	concrete time step $t \leq T$
$\rho_0 : S \to \mathbb{R}^+$	a distribution of start states
$\pi : S \to A$	policy (Note: could also be defined as $S \times A \to \mathbb{R}$ to highlight non-deterministic behavior)
$\pi^* : S \to A$	optimal policy
$G_t : S \to \mathbb{R}$	discounted sum of rewards from time step t to horizon (also with $S \times A$ possible)
$G_t^{(n)} : S \to \mathbb{R}$	general n-step return
$V^\pi : S \to \mathbb{R}$	state-value function: Expected return starting from a given state following policy π
$V^* : S \to \mathbb{R}$	expected return starting from a given state following the optimal policy
$N : S \to \mathbb{R}$	number of times s was visited (also with $S \times A$ possible)
δ_t	TD (temporal difference) error
λ	weight for TD(λ)
$Q^\pi : S \times A \to \mathbb{R}$	state-action value function; follow π after taking the given action
$Q^* : S \times A \to \mathbb{R}$	state-action value function; follow the optimal π after taking the given action
ϵ	probability to do a random exploration step
α	step size for updating (e.g.) the Q-function
$\mathbb{E}$	expectation
∇	gradient
Δ	difference (e.g., update)
∂	partial derivative
$\hat{\cdot}$	approximation of $\cdot$ (e.g., $\hat{V}$ or $\hat{Q}$)
$\mathbf{x}$	feature vector
$\mathbf{w}$ or θ	weight vector or tensor of function approximator
π_θ	policy with policy network parameterized by θ
$d : S \to \mathbb{R}$	stationary distribution over states
τ or h	Trajectory or history (state, action, reward, state, action, ...)
$b : S \to \mathbb{R}$	baseline (estimator) for a given state
μ	mean
σ	standard deviation
H	entropy