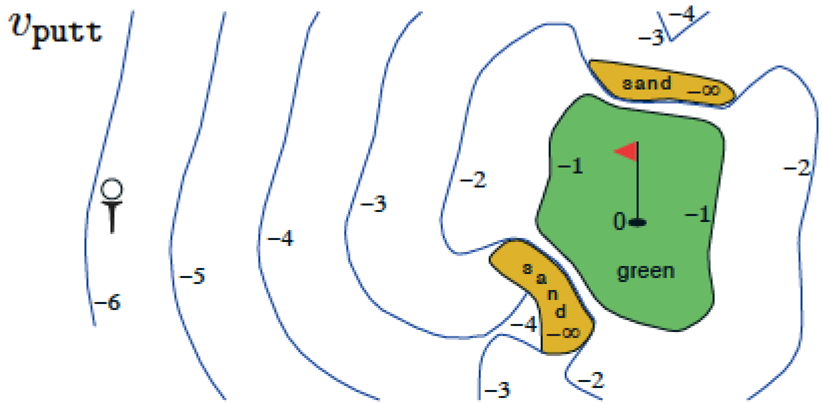
Summary



State-value function for golf-playing agent (Sutton and Barto, 2017)

Policies

- A **deterministic policy** is a mapping $\pi : \mathcal{S} \to \mathcal{A}$. For each state $s \in \mathcal{S}$, it yields the action $a \in \mathcal{A}$ that the agent will choose while in state s.
- A **stochastic policy** is a mapping $\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$. For each state $s \in \mathcal{S}$ and action $a \in \mathcal{A}$, it yields the probability $\pi(a|s)$ that the agent chooses action a while in state s.

State-Value Functions

- The **state-value function** for a policy π is denoted v_π. For each state $s \in \mathcal{S}$, it yields the expected return if the agent starts in state s and then uses the policy to choose its actions for all time steps. That is, $v_\pi(s) \doteq \mathbb{E}_\pi[G_t|S_t = s]$. We refer to $v_\pi(s)$ as the **value of state s under policy π**.
- The notation $\mathbb{E}_\pi[\cdot]$ is borrowed from the suggested textbook, where $\mathbb{E}_\pi[\cdot]$ is defined as the expected value of a random variable, given that the agent follows policy π.

Bellman Equations

- The **Bellman expectation equation for v_π** is: