REINFORCE & Baseline - REINFORCE

Policy fn approximator: Neural Net with 1 hidden layer

W_{in} : n_hidden × n_state_dims
b_{in} : n_hidden

W_{out} : n_actions × n_hidden
b_{out} : n_actions

} parameters

let s_t : n_state_dims $\in S$

$$\pi(s_t) := \text{SOFTMAX}\left(W_{out} \cdot \text{RELU}\left(W_{in} \cdot s_t + b_{in}\right) + b_{out}\right)$$

Prob. dist over $\{1 \ldots n\_actions\}$

* there's also dropout of $p = 0.6$ b/w the layers
n_hidden = 128

lr : 0.01
$\gamma \in [1, 0.99, 0.95]$

Value fn approximator: NN with 1 hidden layer

W_{in} : n_hidden × n_state_dims
b_{in} : n_hidden

W_{out} : 1 × n_hidden
b_{out} : 1

} parameters

let S_t : n_state_dims $\in S$

$$V(S_t) := W_{out} \cdot RELU(W_{in} \cdot S_t + b_{in}) + b_{out}$$

* there's also dropout of $p = 0.6$ b/w the layers
* n_hidden = 128

lr : 0.01
$\gamma \in [1, 0.99, 0.95]$