POLICY PARAMETRIZATION

ARXIV 1908.00261 FOR DETAILS

→ COMPLETE = EVERY STOCHASTIC POLICY = TABULAR

→ $\theta \in \mathbb{R}^{|S||A|}$

INCOMPLETE = ONLY CERTAIN STOCHASTIC POLICIES ARE CONSIDERED, AND WE LOOK FOR THE BEST IN THIS FAMILY

→ $\theta \in \mathbb{R}^{d'}$, with $d' << |S||A|$

HOW TO USE θ TO OBTAIN A POLICY?

1) DIRECT PARAMETRIZATION (ONLY POSSIBLE FOR COMPLETE PARAMETRIZATIONS)

$\pi_\theta(a|s) := \theta_{s,a}$

CONSTRAINTS: $\theta_{s,a} \geq 0$, $\sum_a \theta_{s,a} = 1$ (FOR EVERY s, $(\theta_{s,a})$ LIVES ON A SIMPLEX)

2) SOFTMAX PARAMETRIZATION (COMPLETE AND INCOMPLETE)

$$\pi_\theta(a|s) = \frac{e^{h(s,a,\theta)}}{\sum_b e^{h(s,b,\theta)}}, \text{ FOR PREFERENCES } h(s,a,\theta)$$

→ COMPLETE: $h(s,a,\theta) = \theta_{s,a}$

= UNCONSTRAINED, BECAUSE SOFTMAX GIVES A PROBABILITY DISTRIBUTION

→ LINEAR APPROXIMATOR

→ NEURAL NETWORK

→ MOST USED