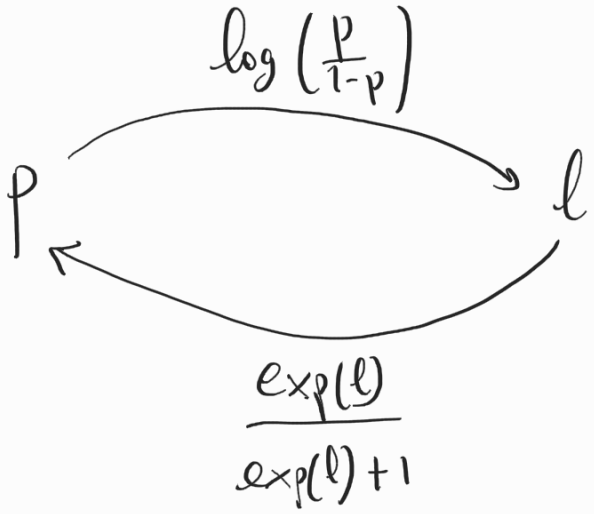
(1) Function approximation

probability → logit

p → $\log\left(\frac{p}{1-p}\right)$

$(0, 1)$

$(-\infty, +\infty)$



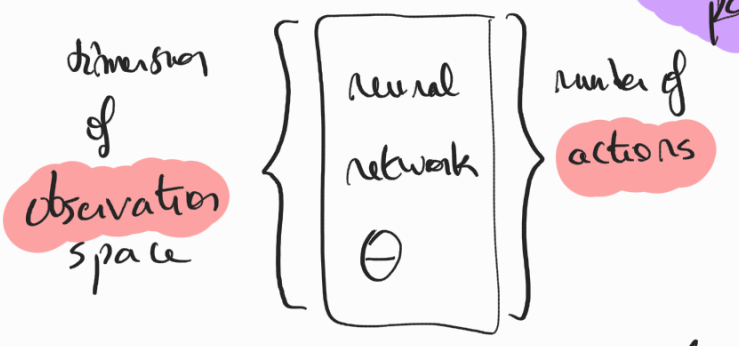
Question: What if S is infinite?

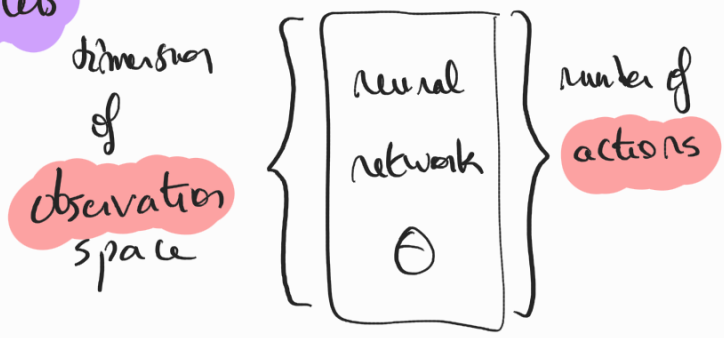
↳ Change of representation

Deterministic polic

Stochastic

θ : set of parameters





$q_\theta : S \times A \to \mathbb{R}$ ← value

$\sigma_\theta(s) = \underset{a \in A}{\operatorname{argmax}}\ q_\theta(s,a)$

$p_\theta : S \times A \to \mathbb{R}$ ← logit

$\sigma_\theta(s) = \text{distribution}(a \mapsto p_\theta(s,a))$

apply softmax here to get probabilities

Update: [NEW = OLD + α(CURRENT − OLD)] ← parameters

↳ includes "gradient ascent algorithms"

Formulation:

$$\min_\theta \mathcal{L}(\theta)$$

↳ loss

Typical approach : **stochastic gradient descent**

Sketch:

(1) **Batch sampling**:
using the current policy, find SETS of trajectories

(2) **Batch update**:

(4) Batch update:

using the batch, update the parameters