s a

d3 ♗b4+

♘f3 ♞c6

0-0 ♗xc3

⋮

$\rightarrow$

Learner

Objective

$$\pi(s) = \texttt{softmax}(g_\theta(s))$$

$$\mathcal{L}(a, \pi(s)) = H(a, \pi(s))$$

Hypothesis space

Convolutional neural net

Optimizer

Stochastic gradient descent

$\rightarrow \quad \pi : s \rightarrow a$

$$\pi^* = \arg\min_{\pi \in \Pi} \sum_{i=1}^{N} \mathcal{L}(\pi(s_i), a_i)$$