Learner

 $\{x^{(i)}, y^{(i)}\}_{i=1}^{N} \to$

Objective



 $\mathcal{L}(\mathbf{y}, f_{ heta}(\mathbf{x})) = H(\mathbf{y}, \mathtt{softmax}(f_{ heta}(\mathbf{x})))$