

$$\mathcal{L}(\theta, X, Y) = \sum_{i=1}^n \sum_{k=1}^C y_{ik} \log p(\theta, x_i, y_i)$$

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

$$\nabla_{\theta} \mathcal{L} = \left\{ \frac{\partial \mathcal{L}}{\partial \theta_0}, \frac{\partial \mathcal{L}}{\partial \theta_1}, \frac{\partial \mathcal{L}}{\partial \theta_2} \dots \frac{\partial \mathcal{L}}{\partial \theta_d} \right\}$$

$$x_i \quad \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array}$$

0	1	0	0
---	---	---	---

$$y_{i,k=0} = 0$$

$$y_{i,k=2} = 0$$

$$y_{i,k=1} = 1$$

$$y_{i,k=3} = 0$$

$$y_i \quad \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}$$

0	0	0	0	0	0	1	0	0
---	---	---	---	---	---	---	---	---

$$x_i$$

one-hot encoding

$$\sum_{k=0}^{C-1} y_k \log p_k$$

~	~	~	~	~	~	~	~	~
*	*	*	*	*	~	~	~	~
0	0	0	0	0	0	1	0	0

$$\log(p)$$

$$y * \log(p)$$