

## Bài 1.

Cho mạng nơron nhân tạo sau

$$\begin{aligned} o_0 &= x, \\ z_\ell &= W_\ell o_{\ell-1} + b_\ell, \ell = 1, 2, \dots, L, \\ o_\ell &= \sigma_\ell(z_\ell), \ell = 1, 2, \dots, L-1, \\ o &= z_L. \end{aligned}$$

Xác suất dự đoán của mẫu đầu vào  $x$  thuộc về lớp  $c$  được tính bằng hàm softmax

$$P(c|x) = \frac{\exp(o_c)}{\sum_{c'=1}^C \exp(o_{c'})}, \quad c = 1, 2, \dots, C. \quad (1)$$

Với nhãn  $y \in \{1, 2, \dots, C\}$ , giả sử ta có hàm mất mát  $L(o, y)$  là hàm entropy chéo

$$L(o, y) = - \sum_{c=1}^C \mathbb{I}(y=c) \log P(c|x), \quad (2)$$

1. Viết công thức tính đạo hàm của hàm mất mát theo đầu ra  $o_a$ , tức là  $\frac{\partial L}{\partial o_a}$ . Sau đó, viết ở dạng véc-tơ  $\delta_L = \nabla_o L$ .

$$L(o, y) = - \sum_{c=1}^C \mathbb{I}(y=c) \cdot \log P(c|x)$$

$$= -\log P(y|x) = -\log p(y)$$

$$\Rightarrow \frac{\partial L}{\partial o_a} = \frac{\partial L}{\partial p(y)} \cdot \frac{\partial p(y)}{\partial o_a}$$

$$\oplus \frac{\partial L}{\partial p(y)} = \frac{-1}{p(y)}$$

$$\oplus \frac{\partial p(y)}{\partial o_a} = \left( \frac{e^{o_y}}{\sum_{i=1} e^{o_i}} \right)'_{o_a}$$

$$= \frac{(e^{o_y})'_{o_a} \sum_{i=1} e^{o_i} - e^{o_y} \cdot (\sum_{i=1} e^{o_i})'_{o_a}}{\left(\sum_{i=1} e^{o_i}\right)^2}$$

$$= \frac{I(y=a) e^{o_y} \sum_{i=1} e^{o_i} - e^{o_y} \cdot e^{o_a}}{\left(\sum_{i=1} e^{o_i}\right)^2}$$

Via  $a=y$ :

$$\frac{\partial p(y)}{\partial o_a} = \frac{e^{o_y}}{\sum_{i=1} e^{o_i}} \cdot \left(1 - \frac{e^{o_y}}{\sum_{i=1} e^{o_i}}\right)$$

$$= p(y) (1 - p(y))$$

$$\Rightarrow \frac{\partial p(y)}{\partial o_a} = p(y) \cdot (1 - p(y))$$

$$\Rightarrow \frac{\partial L}{\partial o_a} = \frac{-1}{p(y)} \cdot p(y) \cdot (1 - p(y)) = p(y) - 1$$

$$= p(a) - 1$$

Via  $a \neq y$ :

$$\frac{\partial p(y)}{\partial o_a} = \frac{-e^{o_y} \cdot e^{o_a}}{\left(\sum_{i=1} e^{o_i}\right)^2} = -p(y) \cdot p(a)$$

$$\Rightarrow \frac{\partial L}{\partial o_a} = \frac{-1}{p(y)} (-p(y) \cdot p(a)) = p(a)$$

$$\Rightarrow \frac{\partial L}{\partial o_a} = p(a) - I(a=y) = P(a|x) - I(a=y)$$

$$\Rightarrow \delta_L = \nabla_o L = \begin{pmatrix} p_1 \\ \vdots \\ p_c \end{pmatrix} - \begin{pmatrix} I(y=1) \\ \vdots \\ I(y=c) \end{pmatrix}$$

2. Viết công thức tính đạo hàm của hàm mất mát đối với đầu ra thứ  $a$  của tầng thứ  $\ell$ , tức là  $\frac{\partial L}{\partial o_{\ell,a}}$  dựa trên đạo hàm của hàm mất mát đối với đầu ra tầng thứ  $\ell + 1$ . Từ đó, viết ở dạng véc-tơ  $\delta_\ell = \nabla_{o_\ell} L$ .

*backperception*

$$z_{\ell+1} = W_{\ell+1} \cdot o_\ell + b_{\ell+1}$$

$$o_\ell = \phi(z_\ell) ; \quad o_{\ell+1} = \phi_{\ell+1}(z_{\ell+1})$$

$$\frac{\partial L}{\partial o_{\ell,a}} = \sum_{b=1}^{d_{\ell+1}} \frac{\partial L}{\partial z_{\ell+1,b}} \cdot \frac{\partial z_{\ell+1,b}}{\partial o_{\ell,a}}$$

$$z_{\ell+1,b} = \sum_a W_{\ell+1,ba} \cdot o_{\ell,a} + b_{\ell+1,b}$$

$$\Rightarrow \frac{\partial z_{\ell+1,b}}{\partial o_{\ell,a}} = W_{\ell+1,ba}$$

$$\Rightarrow \frac{\partial L}{\partial o_{l,a}} = \sum_{b=1}^{d_{l+1}} \frac{\partial L}{\partial z_{l+1,b}} \cdot W_{l+1,b,a}$$

$$\Rightarrow \delta_l = \nabla_{o_l} L = W_{l+1}^T \cdot \delta_{l+1} \quad (\delta_{l+1} = \nabla_{o_{l+1}} L)$$

3. Viết công thức tính đạo hàm của hàm mất mát đối với trọng số  $W_{\ell,a,b}$  và độ lệch  $b_{\ell,a}$  của tầng thứ  $\ell$ , tức là  $\frac{\partial L}{\partial W_{\ell,a,b}}$  và  $\frac{\partial L}{\partial b_{\ell,a}}$ . Sau đó, viết ở dạng ma trận và véc-tơ

$$\nabla_{W_\ell} L \in \mathbb{R}^{d_\ell \times d_{\ell-1}},$$

$$\nabla_{b_\ell} L \in \mathbb{R}^{d_\ell}.$$

$$\frac{\partial L}{\partial W_{\ell,a,b}} = \frac{\partial L}{\partial z_{\ell,a}} \cdot \frac{\partial z_{\ell,a}}{\partial W_{\ell,a,b}}$$

mà  $z_{\ell,a} = \sum_{j=1}^{d_{\ell-1}} W_{\ell,a,j} \cdot o_{\ell-1,j} + b_{\ell,a}$

$$\frac{\partial z_{\ell,a}}{\partial W_{\ell,a,b}} = o_{\ell-1,b} \quad ; \quad \frac{\partial L}{\partial z_{\ell,a}} = \delta_{\ell,a}$$

$$\Rightarrow \frac{\partial L}{\partial W_{\ell,a,b}} = \delta_{\ell,a} \cdot o_{\ell-1,b}$$

$$\frac{\partial L}{\partial b_{l,a}} = \frac{\partial L}{\partial z_{l,a}} \cdot \frac{\partial z_{l,a}}{\partial b_{l,a}} = \frac{\partial L}{\partial z_{l,a}} \cdot 1 \\ = \delta_{l,a}$$

$$\nabla_{W_l} L = \delta_l \cdot o_{l-1}^T$$

$$(d_l \times d_{l-1}) \quad [d_l \times 1] \quad [1 \times d_{l-1}]$$

$$\nabla_{b_l} L = \delta_l$$

4. Viết công thức cập nhật tham số  $W_\ell$  và  $b_\ell$  sử dụng thuật toán SGD (kích thước lô bằng 1) với tốc độ học là  $\eta$ .

SGD:

$$\theta \leftarrow \theta - \eta \nabla_\theta L \\ \Rightarrow W_\ell \leftarrow W_\ell - \eta \frac{\nabla_{W_\ell} L}{\delta_l \cdot o_{l-1}^T}$$

$$b_\ell \leftarrow b_\ell - \eta \frac{\nabla_{b_\ell} L}{\delta_l}$$

$$\Rightarrow \begin{cases} W_{l,ab} \leftarrow W_{l,ab} - \eta \cdot \delta_{l,a} \cdot o_{l-1,b}^T \\ b_{l,a} \leftarrow b_{l,a} - \eta \delta_{l,a} \end{cases}$$