

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

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

- 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 $\delta_L = \nabla_o L$.

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

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

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

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

$$\oplus \frac{\partial p(y)}{\partial \sigma_a} = \left(\frac{e^{\sigma_y}}{\sum_{i=1} e^{\sigma_i}} \right)'_{\sigma_a}$$

$$= \frac{(e^{\sigma_y})'_{\sigma_a} \sum_{i=1} e^{\sigma_i} - e^{\sigma_y} \cdot \left(\sum_{i=1} e^{\sigma_i} \right)'_{\sigma_a}}{\left(\sum_{i=1} e^{\sigma_i} \right)^2}$$

$$= \frac{\mathbb{I}(y=a) e^{\sigma_y} \sum_{i=1} e^{\sigma_i} - e^{\sigma_y} \cdot e^{\sigma_a}}{\left(\sum_{i=1} e^{\sigma_i} \right)^2}$$

$$\forall a=y: \frac{\partial p(y)}{\partial \sigma_a} = \frac{e^{\sigma_y}}{\sum_{i=1} e^{\sigma_i}} \cdot \left(1 - \frac{e^{\sigma_y}}{\sum_{i=1} e^{\sigma_i}} \right)$$

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

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

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

$$= p(a) - 1$$

$$\forall a \neq y: \frac{\partial p(y)}{\partial \sigma_a} = \frac{-e^{\sigma_y} \cdot e^{\sigma_a}}{\left(\sum_{i=1} e^{\sigma_i} \right)^2} = -p(y) \cdot p(a)$$

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

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

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

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

backperception

$$z_{l+1} = W_{l+1} \cdot o_l + b_{l+1}$$

$$o_l = \sigma_l(z_l) \quad ; \quad o_{l+1} = \sigma_{l+1}(z_{l+1})$$

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

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

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

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

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

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

$$\nabla_{W_l} L \in \mathbb{R}^{d_l \times d_{l-1}},$$

$$\nabla_{b_l} L \in \mathbb{R}^{d_l}.$$

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

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

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

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

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

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

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

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

4. Viết công thức cập nhật tham số W_l và b_l sử dụng thuật toán SGD (kích thước lô bằng 1) với tốc độ học là η .

SGD:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} L$$

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

$$b_l \leftarrow b_l - \eta \frac{\nabla_{b_l} L}{\delta_l}$$

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