

# Softmax

①

$$z = Wx + b$$

$$\hat{y}_i = \sigma(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

$$L = - \sum_i y_i \log(\hat{y}_i) \quad \swarrow \text{softmax cross-entropy}$$

Softmax  $\Rightarrow$  Sigmoid (when there are two classes)

$$\begin{aligned} \hat{y}_0 &= \frac{e^{z_0}}{e^{z_0} + e^{z_1}}, \quad \hat{y}_1 = \frac{e^{z_1}}{e^{z_0} + e^{z_1}} \quad \xrightarrow{x = z_1 - z_0} \frac{e^{z_0} e^{e^x}}{e^{z_0} + e^{z_0} e^x} = \frac{e^x}{1 + e^x} \cdot \frac{e^{-x}}{e^{-x}} \\ &\Rightarrow \frac{1}{1 + e^{-x}} \end{aligned}$$

$$\begin{aligned} x = z_1 - z_0 &= (w_1^T x + b_1) - (w_0^T x + b_0) \\ &= (w_1 - w_0)^T x + (b_1 - b_0) \\ &= w^T x + b \end{aligned}$$

$$P(x=1) = \sigma(w^T x + b)$$

$$\begin{aligned} L &= - \sum_{i=0}^1 y_i \log(\hat{y}_i) = -y_0 \log(\hat{y}_0) - y_1 \log(\hat{y}_1) \\ &= -\left( \underset{\uparrow y_1}{y} \log(\hat{y}) + (1-y) \log(1-\hat{y}) \right) \end{aligned}$$

# Sigmoid - Binary Cross-Entropy $\frac{dL}{dw} + \frac{dL}{db}$

(2)

$$L = -[y \log \hat{y} + (1-y) \log (1-\hat{y})]$$

$$\begin{aligned} \frac{dL}{d\hat{y}} &= -\left[ y \cdot \frac{1}{\hat{y}} + (1-y) \left( \frac{-1}{1-\hat{y}} \right) \right] \\ &= -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}} \end{aligned}$$

$$\begin{aligned} \frac{d\hat{y}}{dz} \Rightarrow \hat{y} &= \frac{1}{1+e^{-z}} = (1+e^{-z})^{-1} \\ \frac{d}{dz} &= -(1+e^{-z})^{-2} \cdot (-e^{-z}) \\ &= \frac{e^{-z}}{(1+e^{-z})^2} \Rightarrow \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} \\ &\quad \hat{y} \cdot (1-\hat{y}) \end{aligned}$$

$$\begin{aligned} 1-\hat{y} &= 1 - \frac{1}{1+e^{-z}} \\ \frac{1 \cdot e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}} &= \frac{e^{-z}}{1+e^{-z}} \end{aligned}$$

$$\begin{aligned} \frac{dL}{dz} &= \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{dz} = \left( -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}} \right) (\hat{y} \cdot (1-\hat{y})) \\ &= -y(1-\hat{y}) + (1-y)\hat{y} \\ &= -y + y\hat{y} + \hat{y} - y\hat{y} \\ &= \hat{y} - y \end{aligned}$$

(3)

$$\frac{dz}{dw} = x$$

$$\frac{dL}{dw} = \frac{dL}{dz} \cdot \frac{dz}{dw} = (\hat{y} - y)x$$

$$\frac{dz}{db} = 1$$

$$\frac{dL}{dw} = \frac{dL}{dz} \cdot \frac{dz}{db} = (\hat{y} - y)$$

Softmax + Cross Entropy  $\frac{dL}{dw} + \frac{dL}{db}$

(4)

$$\frac{dL}{d\hat{y}_k} = -\frac{y_k}{\hat{y}_k}$$

$$\frac{d\hat{y}_k}{dz_k} = \frac{d}{dz_k} \left( \frac{e^{z_k}}{\sum_j e^{z_j}} \right) = \frac{e^{z_k} \cdot \sum_j e^{z_j} - e^{z_k} \cdot e^{z_k}}{(\sum_j e^{z_j})^2}$$

$k=m$

$$= \frac{e^{z_k} \left( \sum_j e^{z_j} - e^{z_k} \right)}{(\sum_j e^{z_j})^2}$$

$$= \frac{\hat{y}_k \sum_j e^{z_j} \left( \sum_j e^{z_j} - \hat{y}_k \sum_j e^{z_j} \right)}{(\sum_j e^{z_j})^2}$$

$$= \frac{\hat{y}_k \left( \sum_j e^{z_j} \right)^2 (1 - \hat{y}_k)}{(\sum_j e^{z_j})^2}$$

$$= \hat{y}_k (1 - \hat{y}_k)$$

(5)

$$\begin{aligned} \frac{d\hat{y}_k}{dz_m} \frac{d}{dz_m} \frac{e^{z_k}}{\sum_j e^{z_j}} &= \frac{0 \cdot \sum_j e^{z_j} - e^{z_k} e^{z_m}}{(\sum_j e^{z_j})^2} \\ k \neq m &= \frac{-e^{z_k} e^{z_m}}{(\sum_j e^{z_j})^2} = - \left( \frac{e^{z_k}}{\sum_j e^{z_j}} \right) \left( \frac{e^{z_m}}{\sum_j e^{z_j}} \right) \\ &= -\hat{y}_k \hat{y}_m \end{aligned}$$

$$\begin{aligned} \frac{dL}{dz_m} &= \sum_k \frac{dL}{d\hat{y}_k} \frac{d\hat{y}_k}{dz_m} = -\frac{y_m}{\hat{y}_m} \left( \sum_{n=k}^n \hat{y}_n \right) (1 - \hat{y}_m) + \sum_{k \neq m} \frac{-y_k}{\hat{y}_k} (-\hat{y}_m \hat{y}_k) \\ &= -y_m (1 - \hat{y}_m) + \sum_{k \neq m} y_k \hat{y}_m \\ &= -y_m (1 - \hat{y}_m) + \hat{y}_m \sum_{k \neq m} y_k \\ &= -y_m (1 - \hat{y}_m) + \hat{y}_m (1 - y_m) \\ &= -y_m + y_m \hat{y}_m + \hat{y}_m - y_m \hat{y}_m \\ &= \hat{y}_m - y_m \end{aligned}$$

$$\frac{dZ_m}{dw} = x \quad \frac{dZ_m}{db} = 1$$

$$\frac{dL}{dw} = \frac{dL}{dZ_m} \frac{dZ_m}{dw} = (\hat{y}_m - y_m) x \quad \frac{dL}{db} = (\hat{y}_m - y_m)$$