Derivative Chain Rule with Summation of Terms

Thursday, December 9, 2021 12:09 PM

Let $f(t) = \sigma\left(\begin{bmatrix} w_\alpha^1(t) & w_\beta^1(t) \end{bmatrix} \begin{bmatrix} x_\alpha^1 \\ x_\beta^1 \end{bmatrix}\right) + \sigma\left(\begin{bmatrix} w_\alpha^2(t) + w_\beta^2(t) \end{bmatrix} \begin{bmatrix} x_\alpha^2 \\ x_\beta^2 \end{bmatrix}\right) + \ldots\ldots$

$$f(t) = \sum_{r=1}^{m} \sigma\left(\begin{bmatrix} w_\alpha^r(t) & w_\beta^r(t) \end{bmatrix} \begin{bmatrix} x_\alpha^r \\ x_\beta^r \end{bmatrix}\right)$$

$$= \sum_{r=1}^{m} \sigma\left(w^r(t)^T x^r\right)$$

$$= \sum_{r=1}^{m} g\left(w^r(t)\right)$$

$$\frac{df}{dt} = \sum_{r=1}^{m} \left\langle \frac{dg(w^r(t))}{dw^r(t)}, \frac{dw^r(t)}{dt} \right\rangle$$

This is the general rule.

Let's look at an example.

$$f(t) = \sigma\left(\begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + \sigma\left(\begin{bmatrix} w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}\right)$$

$$= \sigma\left(w^{(1)^T} x^{(1)}\right) + \sigma\left(w^{(2)^T} x^{(2)}\right)$$

$w_1 = t \qquad w_3 = t+1$

$w_2 = t^2 \qquad w_4 = t^3$

$$f(t) = \sigma\left(\begin{bmatrix} t & t^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + \sigma\left(\begin{bmatrix} (t+1) & t^3 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}\right)$$

$$= \sigma\left(x_1 t + x_2 t^2\right) + \sigma\left(x_3(t+1) + x_4 t^3\right)$$

$$\frac{df}{dt} = (x_1 + 2x_2 t)\,\mathbb{I}\left\{\begin{bmatrix} w_1 & w_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} > 0\right\} + (x_3 + 3x_4 t^2)\,\mathbb{I}\left\{\begin{bmatrix} w_3 & w_4 \end{bmatrix}\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} > 0\right\}$$

$$= \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mathbb{I}\left\{\begin{bmatrix} w_1 & w_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} > 0\right\}, \begin{bmatrix} 1 \\ 2t \end{bmatrix} \right\rangle + \left\langle \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} \mathbb{I}\left\{\begin{bmatrix} w_3 & w_4 \end{bmatrix}\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} > 0\right\}, \begin{bmatrix} 1 \\ 3t \end{bmatrix} \right\rangle$$

$$= \left\langle \frac{df}{dw^{(1)}}, \frac{dw^{(1)}}{dt} \right\rangle + \left\langle \frac{df}{dw^{(2)}}, \frac{dw^{(2)}}{dt} \right\rangle$$

$$= \sum_{r=1}^{2} \left\langle \frac{df}{dw^{(r)}}, \frac{dw^{(r)}}{dt} \right\rangle$$