Def. The standard (unit) softmax function is

$$\sigma : \mathbb{R}^n \longrightarrow (0,1)^n$$

where $$\sigma_i(z) = \frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}} .$$

Note. $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$, and $\sigma_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ is just the i^{th} component function (where $i = 1, \ldots, n$).

Rmk. The Jacobian matrix of σ is just the $n \times n$ matrix of partial derivatives:

$$\begin{pmatrix} \frac{\partial \sigma_1}{z_1} & \cdots & \frac{\partial \sigma_1}{z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_n}{z_1} & \cdots & \frac{\partial \sigma_n}{z_n} \end{pmatrix}$$

So, we just have to find $\frac{\partial \sigma_i}{\partial z_j}$ for each $i, j \in \{1, \ldots, n\}$.

We have to use the quotient rule!

Quotient Rule. For a function $q : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $q(x) = \frac{f(x)}{g(x)}$, $q'(x) = \frac{g(x) \cdot f'(x) - f(x) \cdot g'(x)}{[g(x)]^2}$.

Case #1 : $i = j$.

In this situation, $f(z_i) = e^{z_i}$ and $g(z_i) = \sum_{k=1}^{n} e^{z_k}$. Note, $f'(z_i) = e^{z_i} = g'(z_i)$. So,

$$\frac{\partial \sigma_i}{z_i} = \frac{\left[\sum_{k=1}^{n} e^{z_k}\right] \cdot e^{z_i} - e^{z_i} \cdot e^{z_i}}{\left[\sum_{k=1}^{n} e^{z_k}\right]^2}$$

$$= \left(\frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}\right) \cdot \left(\frac{\left[\sum_{k=1}^{n} e^{z_k}\right] - e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}\right)$$

$$= \sigma_i(z) \cdot \left[1 - \sigma_i(z)\right].$$

Case #2 : $i \neq j$.

In this situation, $f(z_j) = e^{z_i}$ and $g(z_j) = \sum_{k=1}^{n} e^{z_k}$. Note, $f'(z_j) = 0$ and $g'(z_j) = e^{z_j}$. So,

$$\frac{\partial \sigma_i}{z_j} = \frac{\left[\sum_{k=1}^{n} e^{z_k}\right] \cdot 0 - e^{z_i} \cdot e^{z_j}}{\left[\sum_{k=1}^{n} e^{z_k}\right]^2}$$

$$= -\left(\frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}}\right) \cdot \left(\frac{e^{z_j}}{\sum_{k=1}^{n} e^{z_k}}\right)$$

$$= -\sigma_i(z) \cdot \sigma_j(z).$$

In summary, $\boxed{\dfrac{\partial \sigma_i}{\partial z_j}} = \begin{cases} \sigma_i(z) \cdot [1 - \sigma_i(z)], & \text{if } i = j \ ; \\ -\sigma_i(z) \cdot \sigma_j(z), & \text{if } i \neq j \ . \end{cases}$

The (i,j)th entry of the Jacobian matrix of σ.