DERIVATION OF SIGMOID FUNCTION

$$\boxed{Sigmoid(x) = \frac{1}{1+e^{-x}}}$$

With the function defined, we can take the derivative with respect to the input, x

$$\frac{d}{dx}S(x) = \frac{d}{dx}\frac{1}{1+e^{-x}}$$

the derivative of a function f(x)with a numerator and denominator can be expressed as:

$$\frac{d}{dx}f = \frac{(denominator * \frac{d}{dx}numerator) - (numerator * \frac{d}{dx}denominator)}{denominator^2}$$

According to the chain rule, the derivative of $f(a^x)$ is $\frac{df}{dx} = \frac{df}{da}\frac{da}{dx}$. Using the chain rule on the denominator, we get $\frac{d(1+e^{-x})}{dx} = -e^{-x}$:

$$\frac{d}{dx}S(x) = \frac{(1+e^{-x})(0) - (1)(-e^{-x})}{(1+e^{-x})^2}$$

Now, we can use the quotient rule to take the derivative:

$$\frac{d}{dx}S(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\frac{d}{dx}S(x) = \frac{1-1+e^{-x}}{(1+e^{-x})^2}$$

$$\frac{d}{dx}S(x) = \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2}$$

$$\frac{d}{dx}S(x) = \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2}$$

$$\frac{d}{dx}S(x) = \frac{1}{(1+e^{-x})}(1 - \frac{1}{1+e^{-x}})$$

$$\frac{d}{dx}S(x) = S(x)(1 - S(x))$$

Derivative of hyperbolic tangent function : <u>*$tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$*</u>

$h(x) = f(x) / g(x)$

$d(h(x)) / dx = (f'(x).g(x) - g'(x).f(x)) / g(x)^2$

or $d(h(x)) / dx = ((df(x)/dx).g(x) - (dg(x)/dx).f(x)) / g(x)^2$

So, we can adapt this rule for hyperbolic tangent function. Because we know that tangent function is quotient of sine and cosine functions.

$tanh(x) = sinh(x) / cosh(x)$

$d(tanh(x))/dx = ((d(sinh(x))/dx).cosh(x) - (d(cosh(x))/dx).sinh(x)) / (cosh(x))^2$

Let's calculate the derivative of sinh(x) and cosh(x)

$sinh(x) = (e^x - e^{-x}) / 2$

$cosh(x) = (e^x + e^{-x}) / 2$

$d(sinh(x))/dx= d ((e^x - e^{-x}) / 2) / dx = d ((e^x/2) - (e^{-x}/2)) / dx = d(e^x/2)/dx - d(e^{-x}/2)/dx = (1/2).(d(e^x)/dx) - (1/2).(d(e^{-x})/dx) = (1/2).e^x - (1/2).e^{-x}.(-1) = (1/2).e^x + (1/2).e^{-x}= (e^x + e^{-x})/2 = cosh(x)$

$d(cosh(x))/dx = d((e^x + e^{-x}) / 2)/dx = d((e^x/2+ e^{-x}/2)/dx = d(e^x/2)/dx + d(e^{-x}/2)/dx = (1/2).d(e^x)/dx + (1/2).d(e^{-x})/dx = (1/2).e^x + (1/2).e^{-x}.(-1) = (1/2).e^x - (1/2).e^{-x} =(e^x-e^{-x})/2 = sinh(x)$

Let's back to calculation of tanh function

$d(tanh(x))/dx = ((d(sinh(x))/dx).cosh(x) - (d(cosh(x))/dx).sinh(x)) / (cosh(x))^2$

$d(tanh(x))/dx = (cosh(x).cosh(x) - sinh(x).sinh(x)) / (cosh(x))^2$

$d(tanh(x))/dx = ((cosh(x))^2 - (sinh(x))^2) / (cosh(x))^2$

$d(tanh(x))/dx = 1 - (sinh(x))^2/(cosh(x))^2 = 1 - (sinh(x)/cosh(x))^2$

$d(tanh(x))/dx = 1 - (tanh(x))^2$

To sum up, hyperbolic tangent function and its derivative are demonstrated as following formulas:

$$f(x) = (e^x - e^{-x}) / (e^x + e^{-x})$$

$$d(f(x))/dx = 1 - (f(x))^2$$

ReLU Function defined as :

$$f(x) = \begin{cases} 0 & \text{if } x \leqslant 0 \\ \text{x} & \text{if } x > 0 \end{cases}$$

Derivative Of ReLU (Rectified Linear Units) Function :

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}$$

And undefined in $x=0$
The reason for it being undefined at $x=0$ is that its left- and right derivative are not equal.

Leaky ReLU :

$$R(z) = \begin{Bmatrix} z & z > 0 \\ \alpha z & z <= 0 \end{Bmatrix}$$

Derivative Of Leaky ReLU Function :

$$R'(z) = \begin{Bmatrix} 1 & z > 0 \\ \alpha & z < 0 \end{Bmatrix}$$

Derivative Of Softmax Function :

$$S_j = \frac{e^{a_j}}{\sum_{k=1}^{N} e^{a_k}} \qquad \forall j \in 1..N$$

Let's compute D_jS_i for arbitrary *i* and *j*:

$$D_j S_i = \frac{\partial S_i}{\partial a_j} = \frac{\partial \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}}{\partial a_j}$$

We'll be using the quotient rule of derivatives. For $f(x) = \frac{g(x)}{h(x)}$:

$$f'(x) = \frac{g'(x)h(x) - h'(x)g(x)}{[h(x)]^2}$$

In our case, we have :

$$g_i = e^{a_i}$$

$$h_i = \sum_{k=1}^{N} e^{a_k}$$

Going back to our D_jS_i we'll start with the i=j case. Then, using the quotient rule we have:

$$\frac{\partial \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}}{\partial a_j} = \frac{e^{a_i}\Sigma - e^{a_j}e^{a_i}}{\Sigma^2}$$

For simplicity ∑ stands for $\sum_{k=1}^{N} e^{a_k}$. Reordering a bit :

$$\begin{aligned} \frac{\partial \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}}{\partial a_j} &= \frac{e^{a_i}\Sigma - e^{a_j}e^{a_i}}{\Sigma^2} \\ &= \frac{e^{a_i}}{\Sigma}\frac{\Sigma - e^{a_j}}{\Sigma} \\ &= S_i(1 - S_j) \end{aligned}$$

Similarly, we can do the I ≠ j case:

$$\begin{aligned} \frac{\partial \frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}}}{\partial a_j} &= \frac{0 - e^{a_j}e^{a_i}}{\Sigma^2} \\ &= -\frac{e^{a_j}}{\Sigma}\frac{e^{a_i}}{\Sigma} \\ &= -S_jS_i \end{aligned}$$

To Summarize :

$$D_jS_i = \begin{cases} S_i(1 - S_j) & i = j \\ -S_jS_i & i \neq j \end{cases}$$