Given x and y, suppose

$$p = \sigma(x) = \frac{1}{1+e^{-x}}$$

and

$$p_t = \begin{cases} p, & if\ y = 1 \\ 1 - p, & otherwise\ (e.g.\ \mathbf{y = 0}) \end{cases}$$

so $p_t = \frac{1}{1+e^{yx}}$

$$\frac{\partial p_t}{\partial x} = \frac{-1}{(1+e^{yx})^2} * y * e^{yx} = y * p_t * (1 - p_t) = -y * p_t * (p_t - 1) ,$$

because of $\partial \frac{1}{x} = \frac{-1}{x^2}$

==================

Suppose Cross-Entropy loss $CE(p_t) = -\log(p_t)$,

given $\partial lnx = \frac{1}{x}$,

$$\frac{\partial CE(p_t)}{\partial x} = \frac{\partial \mathrm{CE}(p_t)}{\partial p_t} * \frac{\partial p_t}{\partial x} = (-\frac{1}{p_t}) * (-y * p_t * (p_t - 1)) = y * (p_t - 1)$$

==================

Suppose Focal Loss $FL(p_t) = -(1-p_t)^\gamma * \log(p_t)$, where γ is const.

$$\frac{\partial FL(p_t)}{\partial x} = \frac{\partial (1-p_t)^\gamma}{\partial x} * (-\log(p_t)) + (1-p_t)^\gamma * \frac{\partial CE(p_t)}{\partial x}$$

$$= (\gamma * (1-p_t)^{\gamma-1} * \frac{\partial 1 - p_t}{\partial p_t}) * \frac{\partial p_t}{\partial x} * (-\log(p_t)) + (1-p_t)^\gamma * y * (p_t - 1)$$

$$= (\gamma * (1-p_t)^{\gamma-1} * \mathbf{-1}) * (-y * p_t * (p_t - 1)) * (-\log(p_t)) + y * (1 - p_t)^\gamma * (p_t - 1)$$

$$= \gamma * (1-p_t)^\gamma * y * p_t * \log(p_t) + y * (1-p_t)^\gamma * (p_t - 1)$$

$$= \boldsymbol{y * (1-p_t)^\gamma * (\gamma * p_t * \log(p_t) + (p_t - 1))}$$

==================

Suppose $\mathrm{z} = \gamma * x + \beta$, and $p_t^* = \sigma(z)$, so $FL^*(p_t^*) = -\log(p_t^*)/\gamma$,

where γ is const

$$\frac{\partial FL^*(p_t^*)}{\partial x} = -\frac{1}{p_t^*} * \frac{1}{\gamma} * \frac{\partial p_t^*}{\partial z} * \frac{\partial z}{\partial x} = -\frac{1}{p_t^*} * \frac{1}{\gamma} * (-y * p_t^* * (p_t^* - 1)) * \gamma$$

$$= \boldsymbol{y * (p_t^* - 1)}$$