Given $x$ and $y$, suppose

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

and

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

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 $\gamma$ is const.

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

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

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

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

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

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

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

where $\gamma$ is const

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

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

## Softmax

Support input data $X = [x_0, x_1, , , , x_{k-1}], where\ k = 0, .., K-1$

$$p_i = \frac{e^{x_i}}{\sum e^{x_k}}$$

suppose $y_i == 1$

if $i == j$

$$\frac{\partial p_i}{\partial x_i} = \frac{e^{x_i} * \sum e^{x_k} - e^{x_i} * e^{x_i}}{\sum {e^{x_k}}^2} = \frac{e^{x_i}}{\sum e^{x_k}} - \frac{e^{x_i}}{\sum e^{x_k}} * \frac{e^{x_i}}{\sum e^{x_k}} = p_i - p_i * p_i = p_i(1 - p_i)$$

if $i \,! = j$

$$\frac{\partial p_i}{\partial x_j} = \frac{0 - e^{x_i} * e^{x_j}}{\sum {e^{x_k}}^2} = -\frac{e^{x_i}}{\sum e^{x_k}} * \frac{e^{x_j}}{\sum e^{x_k}} = -p_i * p_j$$

## Focal Loss

Suppose $FL(p_t) = -\alpha * (1 - p_t)^\gamma * \log(p_t)$, where $\gamma$ is const

$$\frac{\partial FL}{\partial p_t} = [-\alpha * \gamma * (1 - p_t)^{\gamma - 1} * (-1) * \log(p_t)] + [\frac{-\alpha * (1 - p_t)^\gamma}{p_t}]$$

**if $i == j$**

$$\frac{\partial FL}{\partial x_i} = \frac{\partial FL}{\partial p_i} * \frac{\partial p_i}{\partial x_i}$$

$$= \left([-\alpha * \gamma * (1 - p_i)^{\gamma - 1} * (-1) * \log(p_i)] + \left[\frac{-\alpha * (1 - p_i)^\gamma}{p_i}\right]\right) * p_i(1 - p_i)$$

$$= [\alpha * \gamma * (1 - p_i)^\gamma * p_i * \log(p_i)] + [\alpha * (1 - p_i)^\gamma * (p_i - 1)]$$

$$= [-\alpha * \gamma * (\frac{(1 - p_i)^\gamma}{1 - p_i} * (p_i - 1)) * p_i * \log(p_i)] + [\alpha * (1 - p_i)^\gamma * (p_i - 1)]$$

$$= ([-\alpha * \gamma * \frac{(1 - p_i)^\gamma}{1 - p_i} * p_i * \log(p_i)] + [\alpha * (1 - p_i)^\gamma]) * \boldsymbol{(p_i - 1)}$$

**if $i! = j$**

$$\frac{\partial FL}{\partial x_j} = \frac{\partial FL}{\partial p_i} * \frac{\partial p_i}{\partial x_j}$$

$$= \left([-\alpha * \gamma * (1 - p_i)^{\gamma - 1} * (-1) * \log(p_i)] + \left[\frac{-\alpha * (1 - p_i)^\gamma}{p_i}\right]\right) * (-p_i * p_j)$$

$$= ([-\alpha * \gamma * \frac{(1 - p_i)^\gamma}{1 - p_i} * p_i * \log(p_i)] + [\alpha * (1 - p_i)^\gamma]) * \boldsymbol{p_j}$$