Logit Loss

$$L(\theta) = \ln\left(1 + e^{-yf(x,\theta)}\right), y \in \{-1, 1\}, f(x,\theta)$$

求导

$$\begin{aligned}\frac{\partial L_{ll}}{\partial \theta_i} &= \frac{-ye^{-yf(x,\theta)}}{1+e^{-yf(x,\theta)}}\frac{\partial f(x,\theta)}{\partial \theta_i} \\ &= \frac{-y}{1+e^{yf(x,\theta)}}\frac{\partial f(x,\theta)}{\partial \theta_i} \\ &= -y(1-\frac{1}{1+e^{-yf(x,\theta)}})\frac{\partial f(x,\theta)}{\partial \theta_i}\end{aligned} \tag{1}$$

当$y=1$，代入（1）

$$\begin{aligned}\frac{\partial L_{ll}}{\partial \theta_i} &= -(1-\frac{1}{1+e^{-f(x,\theta)}})\frac{\partial f(x,\theta)}{\partial \theta_i} \\ &= \frac{-e^{-f(x,\theta)}}{1+e^{-f(x,\theta)}}\frac{\partial f(x,\theta)}{\partial \theta_i}\end{aligned} \tag{2}$$

当$y=-1$，代入（1）

$$\begin{aligned}\frac{\partial L_{ll}}{\partial \theta_i} &= (1-\frac{1}{1+e^{f(x,\theta)}})\frac{\partial f(x,\theta)}{\partial \theta_i} \\ &= \frac{e^{f(x,\theta)}}{1+e^{f(x,\theta)}}\frac{\partial f(x,\theta)}{\partial \theta_i}\end{aligned} \tag{3}$$

Cross entropy

根据这里的推导，直接给出导数形式, 特别注意$y \in \{0, 1\}$，与Logit Loss不一致。

$$\frac{\partial L_{ce}}{\partial \theta_i} = \left(\frac{1}{1+e^{-f(x,\theta)}} - y\right)\frac{\partial f(x,\theta)}{\partial \theta_i} \tag{4}$$

$y=1$代入（4）

$$\begin{aligned}\frac{\partial L_{ce}}{\partial \theta_i} &= \left(\frac{1}{1+e^{-f(x,\theta)}} - 1\right)\frac{\partial f(x,\theta)}{\partial \theta_i} \\ &= \frac{-e^{-f(x,\theta)}}{1+e^{-f(x,\theta)}}\frac{\partial f(x,\theta)}{\partial \theta_i}\end{aligned} \tag{5}$$

$y=0$代入（4）

$$
\begin{aligned}
\frac{\partial L_{ce}}{\partial \theta_i} &= \frac{1}{1+e^{-f(x,\theta)}}\frac{\partial f(x,\theta)}{\partial \theta_i} \\
&= \frac{e^{f(x,\theta)}}{1+e^{f(x,\theta)}}\frac{\partial f(x,\theta)}{\partial \theta_i}
\end{aligned} \tag{6}
$$

结论

将y代入的目的，是为了理清y的取值范围不同带来的干扰。最后等式(2)与(5)相同，等式(3)与(6)相同，所以Cross Entroy与Logit Loss本质上是相同的损失函数，由于y的取值范围不同，导致形式不一样。