对于逻辑回归模型（二分类）标签只有0和1，单样本$(x^{(i)}, y^{(i)})$的交叉熵损失为

对数损失函数（也叫二元交叉熵损失）$\leftarrow$ $BCELoss = -y^{(i)}\log f_{\vec{w},b}(\vec{x}^{(i)}) + (1-y^{(i)})\log(1-f_{\vec{w},b}(\vec{x}^{(i)}))$

$y^{(i)}$是第i个样本的真实标签，$f_{\vec{w},b}(\vec{x}^{(i)})$代表预测为正样本的概率

而在多分类中（经过softmax层）的单样本$(x^{(i)}, y^{(i)})$的交叉熵损失为：

$$CELoss = \sum_{j=1}^{n} y_j^{(i)} \log(f_{\vec{w},b}(\vec{x}^{(i)})_j) \rightarrow \boxed{\text{多元交叉熵损失}}$$

$y_j^{(i)}$代表的是样本i所属类别j，$f_{\vec{w},b}(\vec{x}^{(i)})_j$表示预测的样本i属于类别j的概率

交叉熵损失函数：交叉熵是一个信息论中的概念，它原来是用来编码长度的
给定两个概率分布p和q，通过q来表示p的交叉熵为

$$H(p,q) = -\sum_x p(x)\log q(x)$$

交叉熵刻画的是两个概率分布之间的距离，p代表正确答案，q代表的是预测值。交叉熵越小，两个概率的分布越接近

交叉熵损失函数的定义为：

$$Loss = -\sum_{i=1}^{n} y_i \log \hat{y}_i \qquad y_i\text{为标签值}, \hat{y}_i\text{为预测值}$$

交叉熵损失函数在二分类的场景中就等价于二元交叉熵损失函数
