Logistic 推导

sigmoid 的一个重要特性是其求导函数可以用其本身来表达。

sigmoid 函数 $y=\frac{1}{1+e^{-z}}$

$$y'=\frac{e^{-z}}{(1+e^{-z})^2}=\frac{1}{1+e^{-z}}\cdot\frac{e^{-z}}{1+e^{-z}}=y(1-y)$$

线性回归模型的公式为 $z=w^Tx+b$

将上式代入sigmoid函数：$y=\frac{1}{1+e^{-(w^Tx+b)}}$

对上式进行转换并取对数：

$$y=1-y\cdot[e^{-(w^Tx+b)}]$$

$$\frac{y}{1-y}=e^{-(w^Tx+b)}$$

$$\ln\frac{y}{1-y}=w^Tx+b \quad \text{式(1)}$$

式(1)即为对数几率回归的模型表达式，如果将y看作是样本作为正例的可能性，那么1-y即为样本作为反例的概率。$\frac{y}{1-y}$也称“几率”(odds)，对几率取对数则得到对数几率。

为了确定模型参数w和b，我们需要推导对数几率回归的损失函数，然后对损失函数进行最小化，得到w和b的估计值。将式(1)中的y视为类后验概率估计$P(y=1|x)$，则对数几率回归模型表达式可重写为：

$$\ln\frac{P(y=1|x)}{P(y=0|x)}=w^Tx+b$$

$$\frac{P(y=1|x)}{P(y=0|x)}=e^{w^Tx+b}=\frac{\frac{e^{w^Tx+b}}{1+e^{w^Tx+b}}}{\frac{1}{1+e^{w^Tx+b}}} \qquad P(y=1|x)=\frac{e^{w^Tx+b}}{1+e^{w^Tx+b}}=\hat{y} \qquad P(y=0|x)=\frac{1}{1+e^{w^Tx+b}}=1-\hat{y}$$

综合上述结果综合可得：$P(y|x)=\hat{y}^{y}(1-\hat{y})^{1-y}$

对上式两边取对数，并取负号：

$$-\ln P(y|x)=-\frac{1}{m}\sum_{i=1}^{m} y\ln\hat{y}+(1-y)\ln(1-\hat{y}) \quad \text{式(2)}$$

式(2)就是经典的交叉熵损失函数，令$L=\ln P(y|x)$，基于L分别对w和b求偏导。

化解式(2)为 $L=-\frac{1}{m}\sum_{i=1}^{m}\left[y\cdot\ln\frac{1}{1+e^{-(w^Tx+b)}}+(1-y)\ln\frac{e^{-(w^Tx+b)}}{1+e^{-(w^Tx+b)}}\right]$

$$=\frac{1}{m}\sum_{i=1}^{m}\left[y(w^Tx+b)-\ln\frac{e^{-w^Tx+b}}{1+e^{-w^Tx+b}}\right]=\frac{1}{m}\sum_{i=1}^{m}\left[y(w^Tx+b)-\ln\left(1-\frac{1}{1+e^{-w^Tx+b}}\right)\right]$$

$$\frac{\partial L}{\partial w}=\frac{1}{m}\sum_{i=1}^{m}\left[y\cdot x-\frac{\cancel{1+e^{-(w^Tx+b)}}}{\cancel{e^{-(w^Tx+b)}}}\cdot\frac{\cancel{e^{-(w^Tx+b)}}\cdot x}{(\cancel{1+e^{-(w^Tx+b)}})^2}\right]=\frac{1}{m}\sum_{i=1}^{m}\left[yx-\frac{1}{1+e^{-w^Tx+b}}\cdot x\right]$$

$$=\frac{1}{m}\sum_{i=1}^{m}(y-\hat{y})x=-\frac{1}{m}\sum_{i=1}^{m}(\hat{y}-y)x$$

$$\frac{\partial L}{\partial b}=\frac{1}{m}\sum_{i=1}^{m}\left[y-\frac{\cancel{1+e^{-(w^Tx+b)}}}{\cancel{e^{-(w^Tx+b)}}}\cdot\frac{\cancel{e^{-(w^Tx+b)}}}{(\cancel{1+e^{-(w^Tx+b)}})^2}\right]=\frac{1}{m}\sum_{i=1}^{m}(y-\hat{y})=-\frac{1}{m}\sum_{i=1}^{m}(\hat{y}-y)$$

基于w和b的梯度下降对交叉熵损失最小化，相应的参数即为模型最优参数。