逻辑回归：

$$z = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b.$$

$$= [w_1, \cdots w_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b = x^T w + b$$

$$\hat{y} = \text{sigmoid}(z) = \frac{1}{1+e^{-z}}$$

损失函数 $J(w) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right]$

来源于：极大似然估计

(1) 确定随机变量分布：

$$\begin{cases} P\{y=1\} = \hat{y} \\ P\{y=0\} = 1-\hat{y} \end{cases} \Rightarrow P(y|\hat{y}) = \begin{cases} \hat{y}, & y=1 \\ 1-\hat{y}, & y=0 \end{cases}$$

$$\Downarrow$$

$$P(y|\hat{y}) = \hat{y}^{y} \cdot (1-\hat{y})^{1-y}$$

$$L = \prod_{i=1}^{m} P(y^{(i)}|\hat{y}^{(i)}) = \prod_{i=1}^{m} \hat{y}^{(i)\,y^{(i)}} (1-\hat{y}^{(i)})^{1-y^{(i)}}$$

$$\log L = \sum_{i=1}^{m}\left[y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})\right]$$

优化目标：$\underset{w^{(i)},b^{(i)}}{\arg\max}\ L \Rightarrow \underset{w^{(i)}b^{(i)}}{\arg\min}\ -\frac{1}{m}L$

梯度计算：

① $$\frac{\partial J(w,b)}{\partial w_j} = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \frac{\partial \log \hat{y}^{(i)}}{\partial w_j} + (1-y^{(i)}) \times \frac{\partial \log(1-\hat{y}^{(i)})}{\partial w_j}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} \times \frac{\partial \log \hat{y}^{(i)}}{\partial \hat{y}^{(i)}} \times \frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} \times \frac{\partial z^{(i)}}{\partial w_j}$$

$$+ (1-y^{(i)}) \times \frac{\partial \log(1-\hat{y}^{(i)})}{\partial (1-\hat{y}^{(i)})} \times \frac{\partial (1-\hat{y}^{(i)})}{\partial \hat{y}^{(i)}} \times \frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} \times \frac{\partial z^{(i)}}{\partial w_j}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} [\, y^{(i)} \times (1-\hat{y}^{(i)}) + (1-y^{(i)})(-\hat{y}^{(i)}) \,] x_j^{(i)}$$

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

② $$\frac{\partial J(w,b)}{\partial b} = -\frac{1}{m}\sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})$$