原始数据：$\left(X_{i}^{'}, y_{i}\right), i=1,\cdots,m$，其中 $X_{i}^{'}=\left(x_{i1}^{'}, x_{i2}^{'},\cdots, x_{in}^{'}\right)\in R^{n}$

训练集：$\left(X_{i}, y_{i}\right), i=1,\cdots,m$，其中 $X_{i}=\left(1, X_{i}^{'}\right)\in R^{n+1}$

在 logistic regression 问题中，logistic 函数表达式如下：

$$h_{\theta}\left(X\right)=g\left(\theta^{T}\cdot X\right)=\frac{1}{1+e^{-\theta^{T}X}}=P\left(y=1\middle|X;\theta\right)$$

现在我们要做的就是让所有训练样本概率最大化，即目标以为

$$\max L\left(w\right)=\prod_{I\left(y_i=1\right)}P\left(y_{i}=1\middle|X_{i};\theta\right)*\prod_{I\left(y_i=0\right)}P\left(y_{i}=0\middle|X_{i};\theta\right)$$

一般的连续的小概率乘法可能会导致浮点下溢 $\log\left(p_{i}\right)+\log\left(p_{j}\right)$ 来代替 $p_{i}\cdot p_{j}$，

其中 $p_{i}, p_{j}>0$

即目标变为

$$\begin{aligned}\min\left(-\log\left(L\left(w\right)\right)\right)&=-\sum_{i=1}^{m}\left[I\left(y_{i}=1\right)\log\left(P\left(y_{i}=1\middle|X_{i};\theta\right)\right)+I\left(y_{i}=1\right)\log\left(P\left(y_{i}=0\middle|X_{i};\theta\right)\right)\right]\\&=-\sum_{i=1}^{m}\left[I\left(y_{i}=1\right)\log\left(P\left(y_{i}=1\middle|X_{i};\theta\right)\right)+I\left(y_{i}=0\right)\log\left(P\left(y_{i}=0\middle|X_{i};\theta\right)\right)\right]\\&=-\sum_{i=1}^{m}\left[y_{i}\log\left(h_{\theta}\left(X_{i}\right)\right)+\left(1-y_{i}\right)\log\left(1-h_{\theta}\left(X_{i}\right)\right)\right]\end{aligned}$$

令 $\min\limits_{\left(X_i, y_i\right), i=1,\cdots,m} J\left(\theta\right)=-\frac{1}{m}\sum_{i=1}^{m}\left[y_{i}\log\left(h_{\theta}\left(X_{i}\right)\right)+\left(1-y_{i}\right)\log\left(1-h_{\theta}\left(X_{i}\right)\right)\right]$

用牛顿法进行求解，先求 hessian 矩阵每个元素 $\frac{\partial J\left(\theta\right)}{\partial\theta_{k}\partial\theta_{m}}$，其中 $k, m\in 1,\cdots,n+1$

$$\frac{\partial J(\theta)}{\partial\theta_k} = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i\frac{1}{h_\theta(X_i)}\cdot h_\theta(X_i)\big(1-h_\theta(X_i)\big)\Big(X_i^{(k)}\Big)+(1-y_i)\frac{-1}{1-h_\theta(X_i)}h_\theta(X_i)\big(1-h_\theta(X_i)\big)\Big(X_i^{(k)}\Big)\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\Big[\big(y_i-h_\theta(X_i)\big)X_i^{(k)}\Big]$$

则 $\frac{\partial J(\theta)}{\partial\theta_k\partial\theta_m} = \frac{1}{m}\sum_{i=1}^{m}\Big[h_\theta(X_i)\big(1-h_\theta(X_i)\big)X_i^{(k)}X_i^{(m)}\Big]$

$$H = \frac{1}{m}\sum_{i=1}^{m}\Big[h_\theta(X_i)\big(1-h_\theta(X_i)\big)X_iX_i^T\Big]$$

其中 $X_i \in R^{n+1}, X_iX_i^T \in R^{(n+1)\times(n+1)}$

其中一阶导函数：

$$\nabla_\theta J = \frac{1}{m}\sum_{i=1}^{m}\big(h_\theta(X_i)-y_i\big)X_i$$

如果采用牛顿法来求解回归方程中的参数，则参数的迭代公式为：

$$\theta^{(t+1)} = \theta^{(t)} - H^{-1}\nabla_\theta J$$