Logistic 回归

利用线性回归函数来进行分类

其中 Sigmoid 函数为:

$$f(z)=\frac{1}{1+e^{-z}}$$

$$\because z=w_0+w_1x_1+w_2x_2\cdots+w_nx_n=w^Tx$$

$$\therefore f(w^Tx)=\frac{1}{1+e^{-w^Tx}}$$

当求出的 $f(x)>0.5$ 时,判断为 1, $f(x)<0.5$ 时,判断为 0.

$$P(Y=1|x)=\frac{e^{w^Tx}}{1+e^{w^Tx}}=\frac{1}{1+e^{-w^Tx}}$$

$$P(Y=0|x)=\frac{1}{1+e^{w^Tx}}=1-P(Y=1|x)$$

下面是利用 sigmoid 来推导其极大似然化模型

$$\because g(z)=\frac{1}{1+e^{-z}}$$

$$z=\theta_0+\theta_1x_1+\theta_2x_2+\cdots+\theta_nx_n=\theta^Tx$$

$$\therefore h_\theta(x)=g(\theta^Tx)=\frac{1}{1+e^{-\theta^Tx}}$$

$$\therefore \begin{cases} P(Y=1|x)=h_\theta(x) \\ P(Y=0|x)=1-h_\theta(x) \end{cases}$$

$$\therefore P(Y|x;\theta)=[h_\theta(x)]^Y[1-h_\theta(x)]^{1-Y}$$

$\therefore$ 似然函数为

$$L(\theta)=\prod_{i=1}^{n}P(y_i|x_i;\theta)=\prod_{i=1}^{n}[h_\theta(x_i)]^{y_i}[1-h_\theta(x_i)]^{1-y_i}$$

其两边取对数,得

$$l(\theta)=\log L(\theta)=\sum_{i=1}^{n}[y_i\log(h_\theta(x_i))+(1-y_i)\log(1-h_\theta(x_i))]$$

对其进行求导:

$$\frac{\partial}{\partial\theta_j}l(\theta)=\sum_{i=1}^{n}[y_i\cdot\frac{1}{h_\theta(x_i)}\cdot\frac{\partial}{\partial\theta_j}h_\theta(x_i)-(1-y_i)\frac{1}{1-h_\theta(x_i)}\frac{\partial}{\partial\theta_j}h_\theta(x_i)]$$

$$=\sum_{i=1}^{n}[y_i\cdot\frac{1}{h_\theta(x_i)}-(1-y_i)\frac{1}{1-h_\theta(x_i)}]\cdot\frac{\partial}{\partial\theta_j}h_\theta(x_i)$$

对于 sigmoid 函数,有一个重要性质:

$$f(x)=\frac{1}{1+e^{-g(x)}}$$

$$\therefore \frac{d}{dx}f(x)=\frac{1}{(1+e^{-g(x)})^2}\cdot e^{-g(x)}\cdot\frac{d}{dx}g(x)=\frac{1}{1+e^{-g(x)}}\cdot\frac{e^{-g(x)}}{1+e^{-g(x)}}\cdot\frac{d}{dx}g(x)=f(x)\cdot(1-f(x))\cdot\frac{d}{dx}g(x)$$

即 $f'(x)=f(x)\cdot(1-f(x))\cdot g'(x)$

$$又\because h_\theta(x)=g(\theta^Tx)=\frac{1}{1+e^{-\theta^Tx}}$$

$$\therefore \frac{\partial}{\partial\theta}h_\theta(x)=\frac{\partial}{\partial\theta}g(\theta^Tx)=\frac{1}{(1+e^{-\theta^Tx})^2}\cdot e^{-\theta^Tx}\cdot x$$

$$=\frac{1}{1+e^{-\theta^Tx}}\cdot\frac{e^{-\theta^Tx}}{1+e^{-\theta^Tx}}\cdot x=h_\theta(x)(1-h_\theta(x))\cdot x$$

$$\therefore \frac{\partial}{\partial\theta_j}l(\theta)=\sum_{i=1}^{n}[y_i(1-h_\theta(x_i))-(1-y_i)h_\theta(x_i)]\,x_i^j$$

$$=\sum_{i=1}^{n}[y_i-h_\theta(x_i)]\,x_i^j$$

对于梯度上升法而言, $\theta_j:=\theta_j+\alpha(\sum_{i=1}^{n}[y_i-h_\theta(x_i)])\cdot x_i^j$

即 $\theta_0:=\theta_0+(Y-h(x))\cdot x_0$ $\longrightarrow$ 利用 x_0 来更新 θ_0, 此时 $x_0=1$