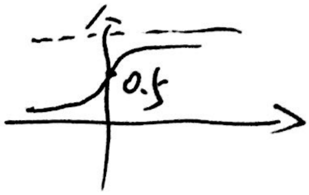
Logistic Model

① 定义概率比 $= \dfrac{\text{发生概率}}{\text{不发生概率}} = \dfrac{p}{1-p}$

② 概率比的对数函数：$logit(p) = \log\left(\dfrac{p}{1-p}\right) \in [0,1]$

③ 反函数：$logistic(z) = sigmod(z) = \dfrac{1}{1+e^{-z}}$



给定测试 $X(x_0, x_1 \cdots x_n)$ 参数 $\theta(\theta_0, \theta_1, \cdots \theta_n)$

$$z = \theta_0 x_0 + \theta_1 x_1 + \cdots\cdots + \theta_n x_n = \Theta^T X$$

$\downarrow$ 因为 $g(z) = \dfrac{1}{1+e^{-z}}$

预测函数 $h_\theta(z) = g(\Theta^T X) = \dfrac{1}{1+e^{-\Theta^T X}}$

④ 对二分类问题
$$\begin{cases} \text{正例 } (y=1): & P(y=1 \mid X;\Theta) = h_\theta(X) \\ \text{反例 } (y=0) & P(y=0 \mid X;\Theta) = 1-h_\theta(X) \end{cases}$$

$\downarrow$

函数合二为一：$P(y \mid X;\Theta) = h_\theta(X)^y [1-h_\theta(X)]^{1-y}$

表示输入向量X的第i个分量

⑤ 目标函数：$L(\Theta) = P(y \mid \overset{x_1, x_2 \cdots x_n}{X}; \overset{\theta_1, \theta_2 \cdots \theta_n}{\Theta}) = \prod_i P(y^{(i)} \mid x^{(i)};\theta) = \prod_i h_\theta(z^{(i)})^{y(i)} [1-h_\theta(z^{(i)})]^{1-y(i)}$

$\downarrow$ 取对数

$$L(\Theta) = \log(L(\Theta)) = \log\left(\prod_i P(y^{(i)} \mid x^{(i)};\theta)\right)$$
$$= \log\left[\prod_i h_\theta(z^{(i)})^{y(i)} [1-h_\theta(z^{(i)})]^{1-y(i)}\right]$$
$$= \sum_{i=1}^{m} \left[y^{(i)} \log h_\theta(z^{(i)}) + (1-y^{(i)}) \log[1-h_\theta(z^{(i)})]\right]$$

求导：∵ $h_\theta(z)=\frac{1}{1+e^{-z}}$ 其中 $z=\Theta^T X=\theta_0x_0+\theta_1x_1+\cdots\theta_mx_m$

$$\frac{\partial L(\theta)}{\partial\theta}=\sum_{i=1}^{m}\left[y^{(i)}\frac{1}{h_\theta(z^{(i)})}-(1-y^{(i)})\frac{1}{1-h_\theta(z^{(i)})}\right]\frac{\partial h_\theta(z^{(i)})}{\partial\theta} \quad <1>$$

而 $$\frac{\partial h_\theta(z)}{\partial\theta}=\frac{\partial}{\partial\theta}\left[\frac{1}{1+e^{-z}}\right]=\frac{e^{-z}}{(1+e^{-z})^2}=\frac{1}{(1+e^{-z})}\left(1-\frac{1}{1+e^{-z}}\right)$$

$$=h_\theta(z)(1-h_\theta(z)) \quad <2>$$

将<2>代入<1>

$$\frac{\partial L(\theta)}{\partial\theta}=\sum_{i=1}^{m}\left[y^{(i)}\frac{1}{h_\theta(z^{(i)})}-(1-y^{(i)})\left(\frac{1}{1-h_\theta(z^{(i)})}\right)\right]h_\theta(z^{(i)})(1-h_\theta(z^{(i)}))\frac{\partial z^{(i)}}{\partial\theta_j}$$

$$=\sum_{i=1}^{m}\left[y^{(i)}(1-h_\theta(z^{(i)}))-(1-y^{(i)})h_\theta(z^{(i)})\right]\cdot x_j$$

$(\because z=\Theta^T X=\theta_0x_0+\theta_1x_1+\cdots+\theta_mx_m)$

$(\therefore \frac{\partial z^{(i)}}{\partial\theta_j}=x_j \ (j=0,1,2,\ldots,m))$

$$\therefore 原式=\sum_{i=1}^{m}\left[y^{(i)}-h_\theta(z^{(i)})\right]x_j$$

⑦ 利用梯度下降法求解目标函数最大值

⇒ 给定前续步长 α 和初始 θ，迭代收敛

<更新规则> $\theta:=\theta+\alpha\nabla_\theta L(\theta)$

而 $\nabla_\theta L(\theta)=\frac{\partial L(\theta)}{\partial\theta_j}=(y^{(i)}-h_\theta(z^{(i)}))x_j^{(i)}$

⑧ repeat until: Converage{

for $i=1$ to m{

$\theta_j:=\theta_j+\alpha[(y^{(i)}-h_\theta(z^{(i)}))x_j^{(i)}]$

}

朴素贝叶斯 Native Bayes

贝叶斯公式(后验概率)

① $\because P(A|B)=\dfrac{P(A\cap B)}{P(B)} \Rightarrow P(A\cap B)=P(A|B)P(B)$

$P(B|A)=\dfrac{P(A\cap B)}{P(A)} \Rightarrow P(A\cap B)=P(B|A)P(A)$

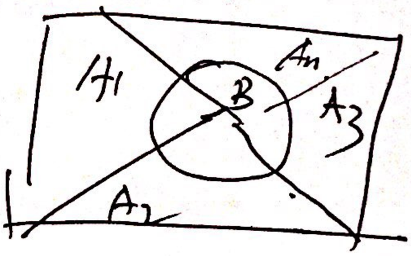
$\therefore P(A|B)P(B)=P(B|A)P(A)$

$\Rightarrow P(A|B)=\dfrac{P(B|A)P(A)}{P(B)}$

后验：$P(A_i|B)=\dfrac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n}P(B|A_j)P(A_j)}$ $i=1,2\cdots n$

发生事件概率　给定条件

② 全概率事件



$P(B)=P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+\cdots+P(B|A_n)P(A_n)$

$=\sum_{i=1}^{n}P(B|A_i)P(A_i)$

③ 根据事件之间的独立性

$P(B|A_i)=P(B=b^1,b^2\cdots b^n|A_i)=\prod_{j=1}^{n}P(B^{(j)}=b^{(j)}|A_i)$

④ 后验 $P(A_i|B)=\dfrac{P(A_i)\prod_{j=1}^{n}P(B^{(j)}=b^{(j)}|A_i)}{\sum_{j=1}^{n}P(B|A_j)P(A_j)=\sum_{j=1}^{n}\left[\prod_{j=1}^{n}P(B^{(j)}=b^{(j)}|A_i)\right]P(A_j)}$

正比于<因为对所有后验概率,它的分母都一致>

⑤ $y=f(x)=P(A_i|B)\propto \arg\max\left\{P(A_i)\prod_{j=1}^{n}P(B^{(j)}=b^{(j)}|A_i)\right\}$