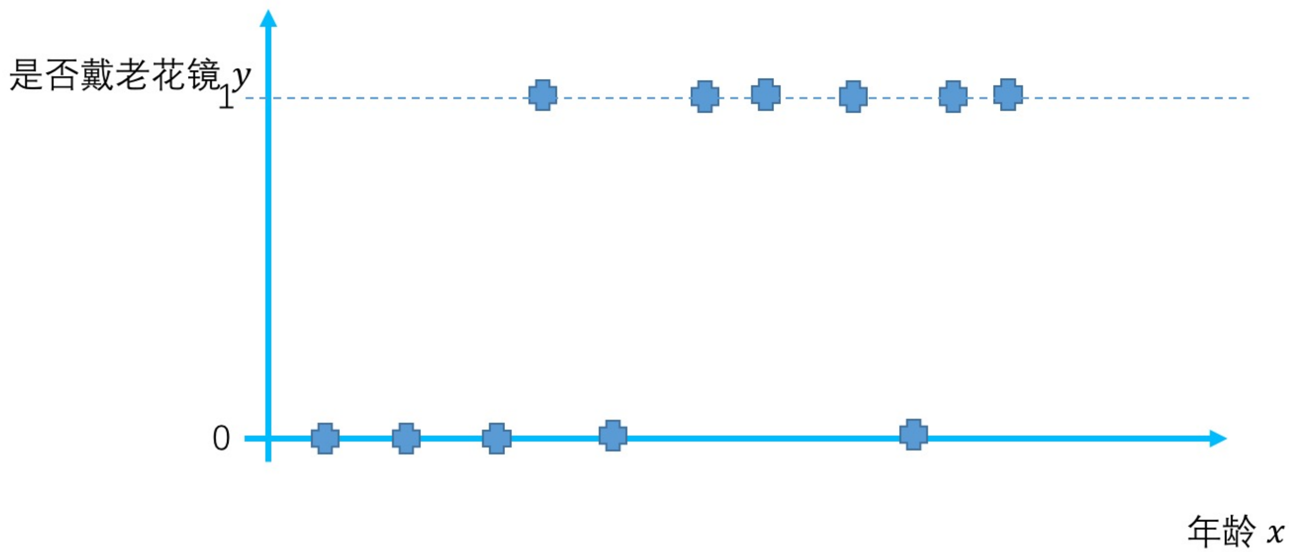
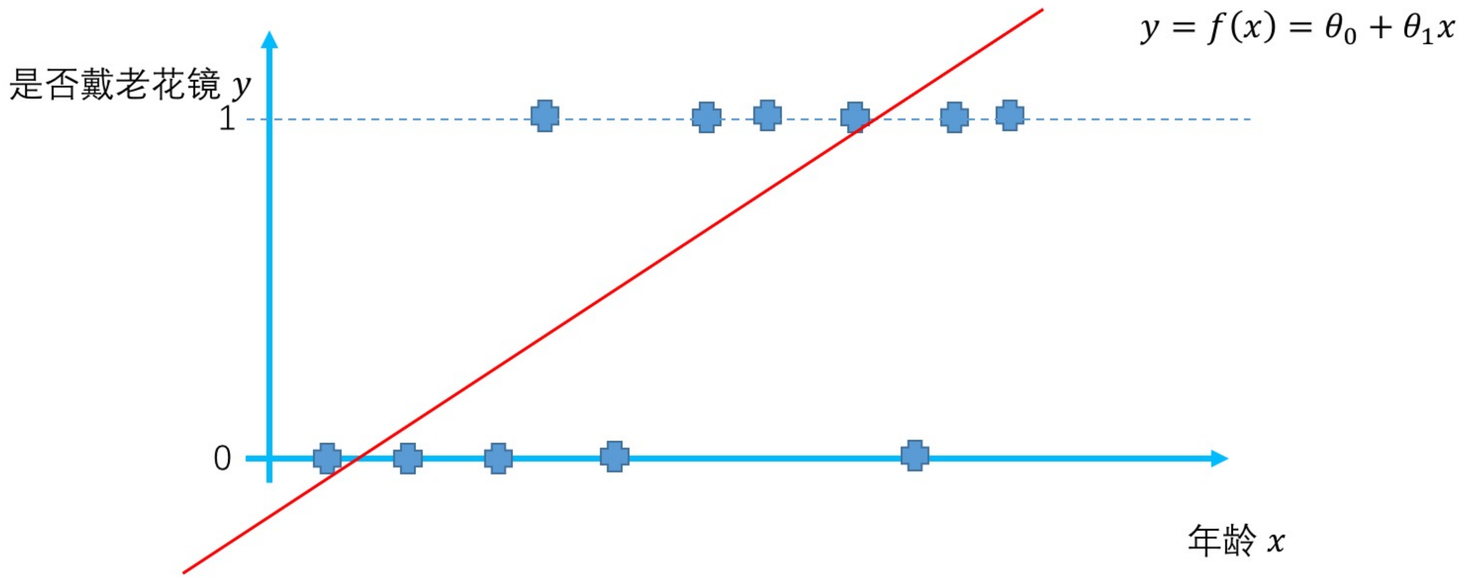
贪心科技
让每个人享受个性化教育服务

逻辑回归

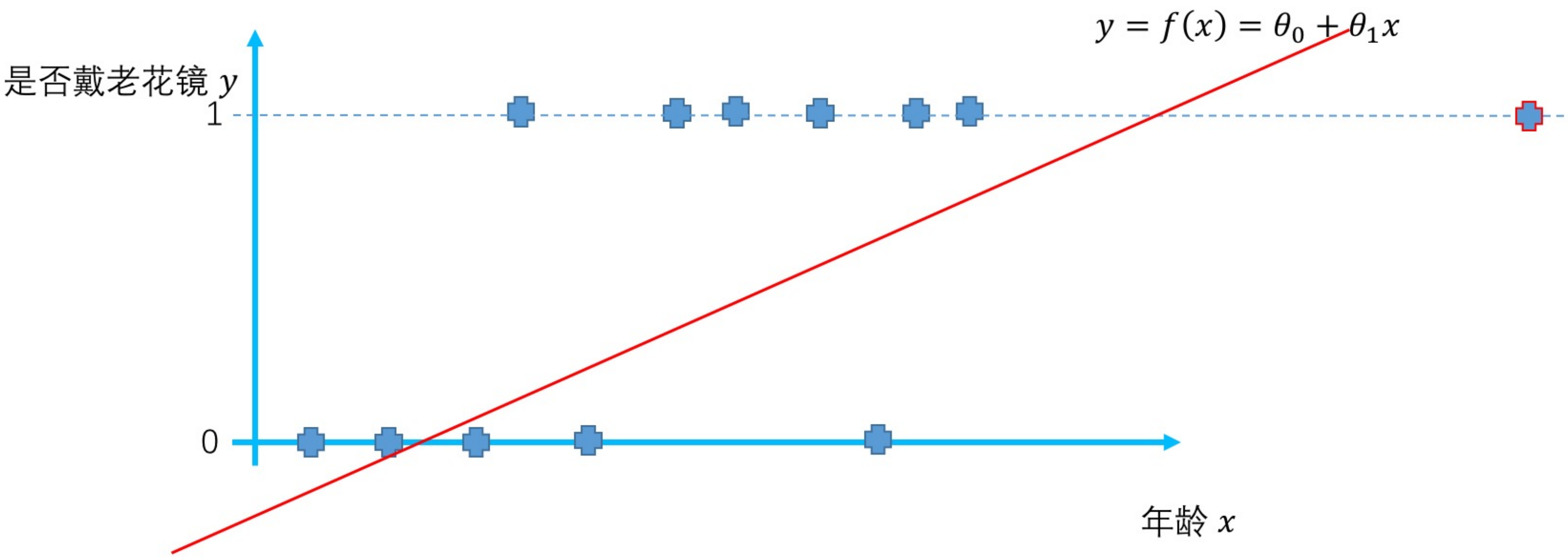
二元分类问题



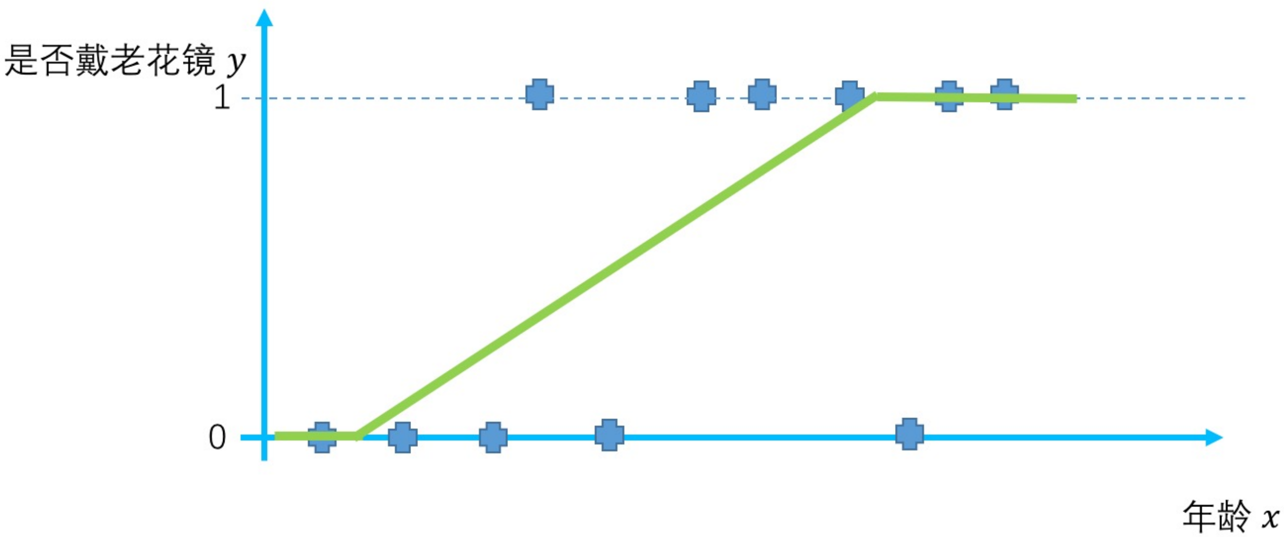
二元分类问题



二元分类问题 (对Outlier敏感)

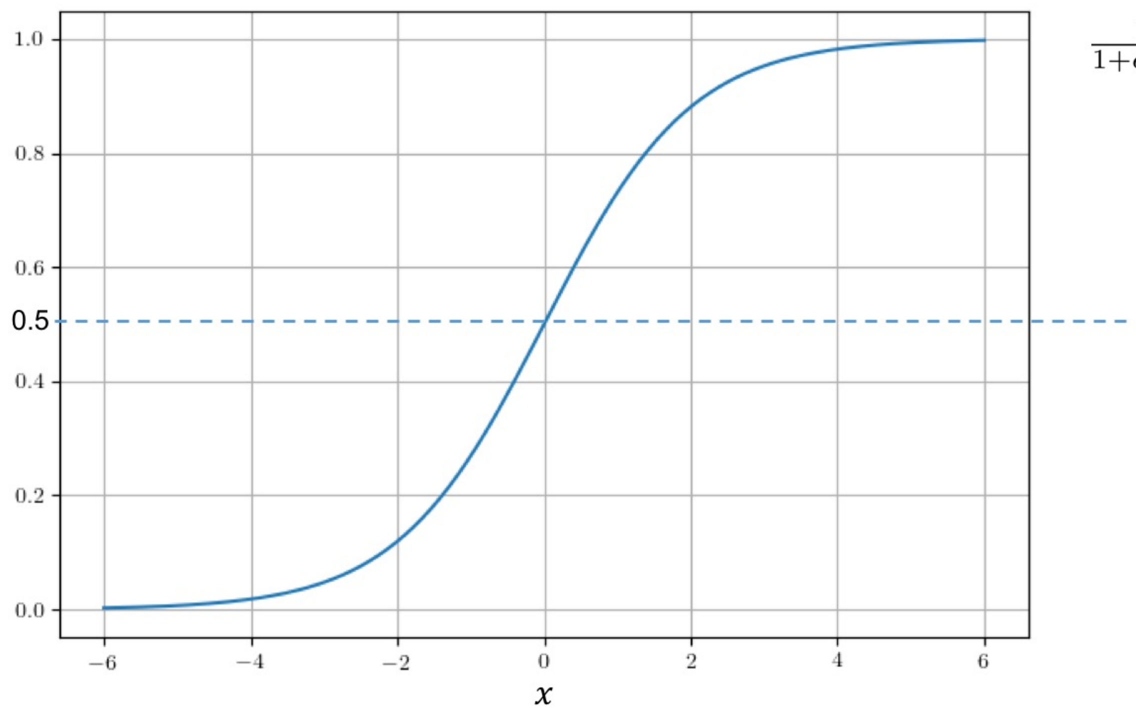


二元分类问题



逻辑函数

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

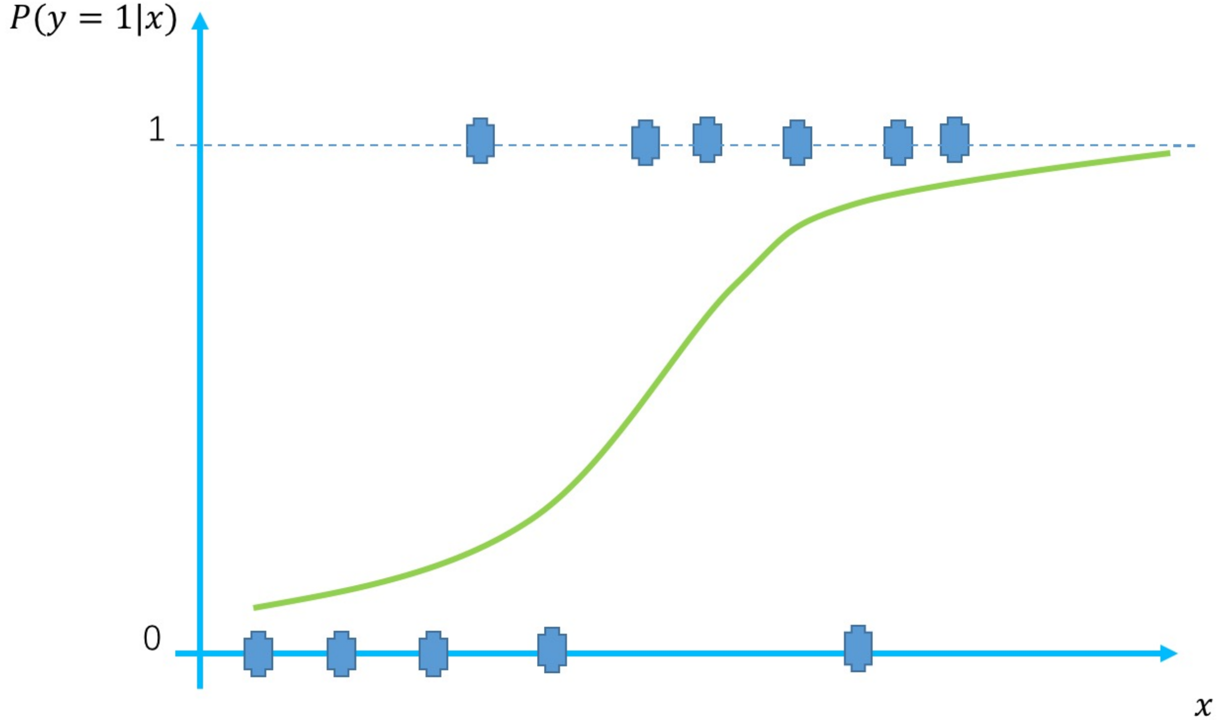


$$\frac{1}{1+e^{-\infty}} = \frac{1}{1+0} = 1$$

$$\frac{1}{1+e^{\infty}} = \frac{1}{1+\infty} = 0$$

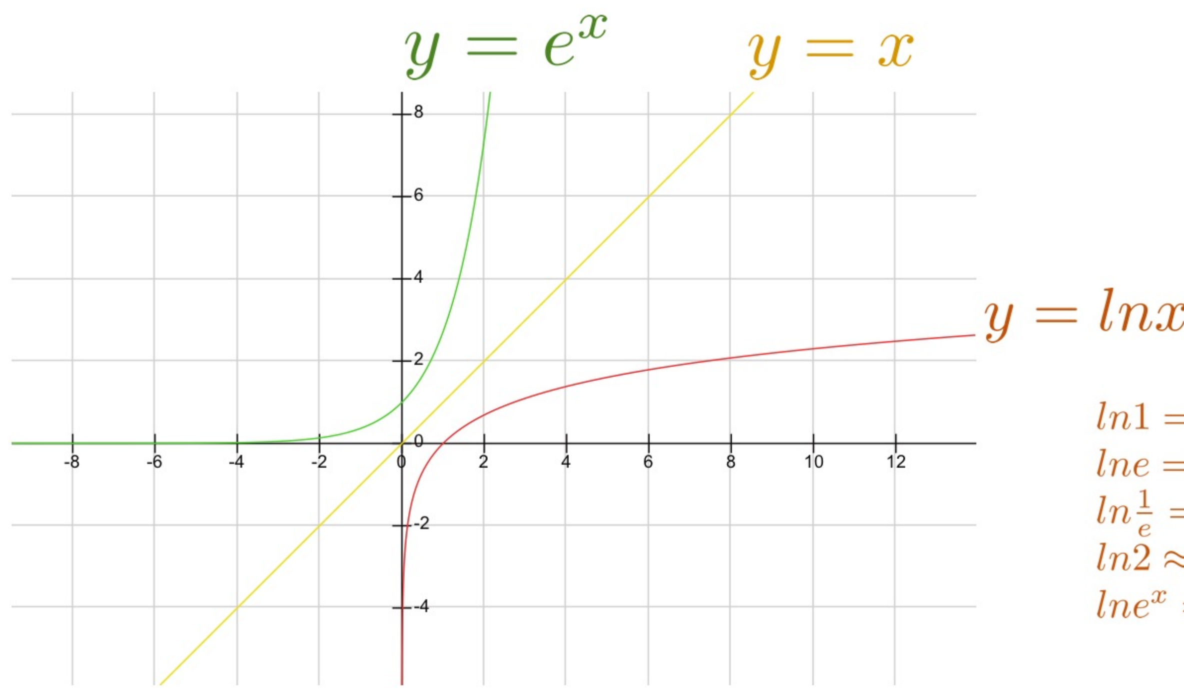
逻辑回归

$$P(y=1|x)=f(x)=\frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$$



指数与对数

$$e^0 = 1$$
$$e^1 = e \approx 2.72$$
$$e^{-1} = \frac{1}{e}$$
$$e^{0.693} \approx 2$$
$$e^{lnx} = x$$



$$ln1 = 0$$
$$lne = 1$$
$$ln\frac{1}{e} = -1$$
$$ln2 \approx 0.693$$
$$lne^x = x$$

逻辑回归

- 解决二元（0/1) 分类的问题
- $P(y=1|x;\theta)=f(x;\theta)=\frac{1.0}{1.0+e^{-\theta^T x}}$
- $\theta^T x=\theta_0+\theta_1 x_1+\theta_2 x_2+\theta_3 x_3+\ \ldots$
- $\theta=[\theta_0,\theta_1,\theta_2,\theta_3,\ldots]$
- $x=[1,x_1,x_2,x_3,\ldots]$

- 当P($y=1|x$)的值大于0.5，输出1； 否则输出0

逻辑回归知识点

类别1的概率	$P=\frac{1}{1+e^{-(\theta^T x)}}$
类别0的概率	$1-P=\frac{1+e^{-\theta^T x}-1}{1+e^{-\theta^T x}}=\frac{1}{1+e^{\theta^T x}}$
类别1与0概率比值	$\frac{P}{1-P}=e^{\theta^T x}$
类别1与0概率比值的 自然对数	$ln\frac{P}{1-P}=\theta^T x$

示例

年龄（x_1）	年收入(x_2)(万元为单位)	是否买车（1表示是，0表示否）
20	3	0
23	7	1
31	10	1
42	13	1
50	7	0
60	5	0
28	8	?

示例

年龄（x_1）	年收入(x_2)(万元为单位)	是否买车（1表示是，0表示否）
20	3	0
23	7	1
31	10	1
42	13	1
50	7	0
60	5	0
28	8	?

$$P(Y=1|x;\theta)=f(x;\theta)=\frac{1}{1+e^{-\theta^\top x}}$$

θ_0 = -0.04, θ_1 = -0.20, θ_2=0.92

$\theta^T x = -0.04 - 0.20 * 28 + 0.92 * 8$ = 1.75

$P(y=1|x) = \frac{1.0}{1.0+ e^{-1.75}}$=0.85

模型训练

年龄（x_1）	年收入(x_2)(万元为单位)	是否买车（1表示是，0表示否）
20	3	0
23	7	1
31	10	1
42	13	1
50	7	0
60	5	0

$$P(y=1|x) = \frac{1.0}{1.0 + e^{-\theta^T x}}$$

如何根据左边的训练数据得到系数的值：

θ_0= -0.04, θ_1 = -0.20, θ_2=0.92

贪心科技 GREEDY TECHNOLOGY

贪 心 科 技　让 每 个 人 享 受 个 性 化 教 育 服 务

THANKS

贪心学院讲师：袁源

Appendix

损失函数

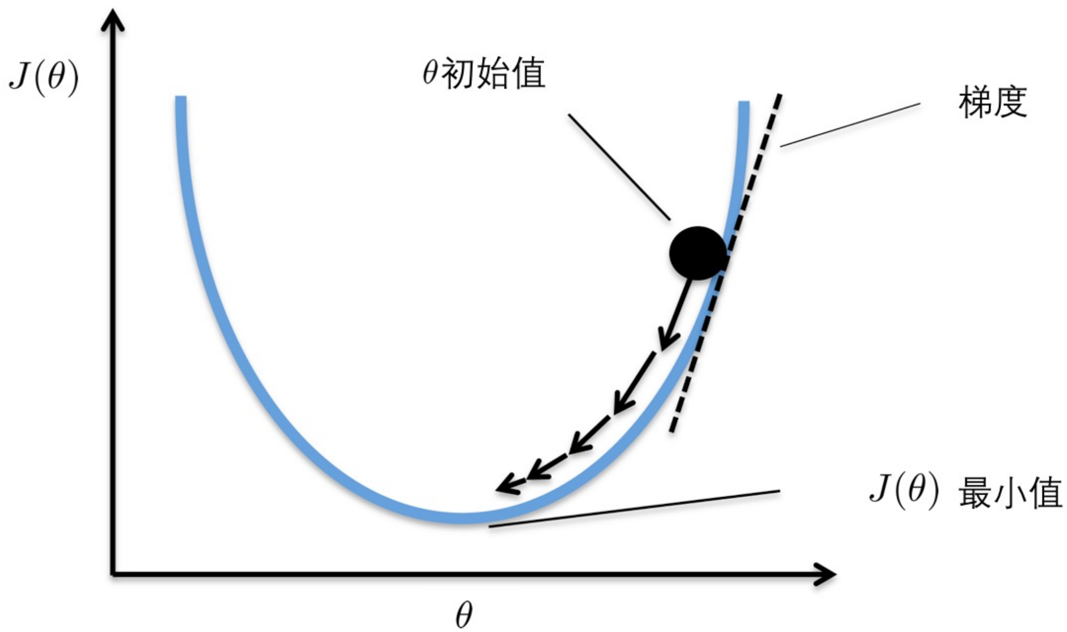
损失函数

$$P(Y=1|x;\theta)=f(x;\theta)=\frac{1}{1+e^{-\theta^\top x}}$$

$$J(\theta)=-\sum_{i=1}^{N}y^{(i)}ln(P(Y=1|X=x^{(i)};\theta))+(1-y^{(i)})ln(1-P(Y=1|X=x^{(i)};\theta))$$

$$\nabla_\theta J(\theta)=\sum_i x^{(i)}(f(x^{(i)};\theta)-y^{(i)})$$

损失函数



梯度下降法

梯度下降法

$$f(x;\theta) = \frac{1}{1+\ e^{-\theta^T x}}$$

$$\nabla_\theta J(\theta) = \sum_i x^{(i)}(f(x^{(i)};\theta) - y^{(i)})$$

特征的值

$$\theta = \theta - \alpha \nabla_\theta J(\theta) = \theta - \alpha \sum_i x^{(i)}(f(x^{(i)};\theta) - y^{(i)})$$

新的值　　旧的值　　学习率　　预测类别1的概率　　已知类别的值

系数的意义

$$P(Y=1|x;\theta)=f(x;\theta)=\frac{1}{1+\ e^{-\theta^T x}}$$

概率比值$odds=\frac{p}{1-p}=e^{\theta^T x}$

系数θ_j意味着，假设原来的$odds$为λ_1，若对应的特征x_j增加1，假设新的$odds$为λ_2， 那么 $\frac{\lambda_2}{\lambda_1}\equiv e^{\theta_j}$

模型训练

年龄（x_1）	年收入(x_2)(万元为单位)	是否买车（1表示是，0表示否）
20	3	0
23	7	1
31	10	1
42	13	1
50	7	0
60	5	0

$$P(Y=1|x;\theta) = \frac{1.0}{1.0 + e^{-\theta^T x}}$$

如何根据左边的训练数据得到系数θ的值：
θ_0 = -0.04, θ_1 = -0.20, θ_2=0.92

系数$\theta_2 = 0.92$ 意味着，如果年收入增加1万，一个人买车和不买车的概率的比值与之前的比值相比较，增加$e^{0.92}$=2.5倍

系数$\theta_1 = -0.20$ 意味着，如果年龄增加1岁，一个人买车和不买车的概率的比值与之前的比值比较降低$e^{-0.20}$=0.82倍

应用场景

- 垃圾邮件分类
- 广告点击预测
- 医疗效果预测

似然函数

$$P(Y=1|x;\theta)=f(x;\theta)=\frac{1}{1+e^{-\theta^\top x}}$$

$$P(Y=0|x;\theta)=1-P(Y=1|x;\theta)$$

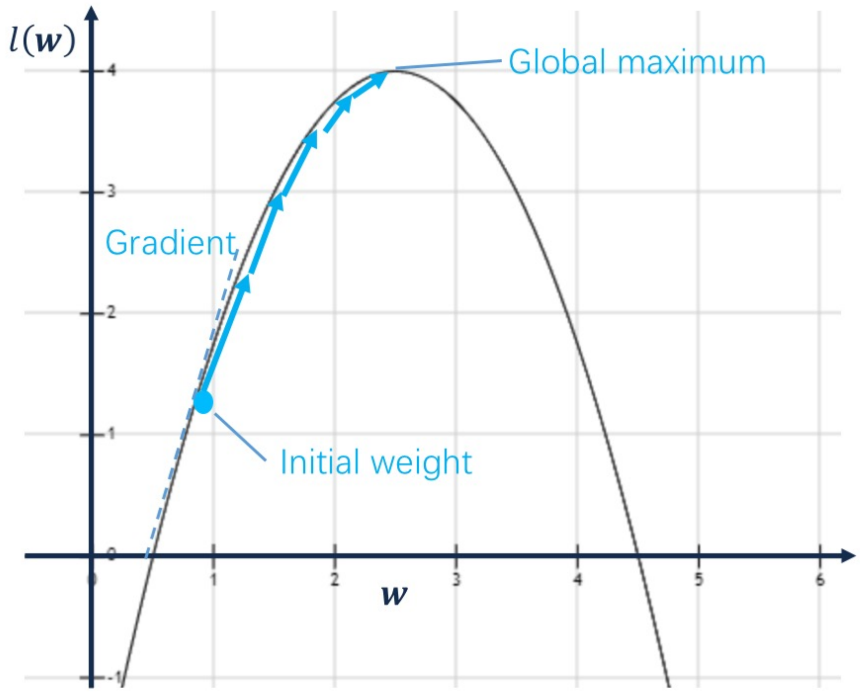
$$L(\theta)=\prod_{i\in\{1,\ldots,N\},y^{(i)}=1}P(Y=1|X=x^{(i)};\theta)\cdot\prod_{i\in\{1,\ldots,N\},y^{(i)}=0}P(Y=0|X=x^{(i)};\theta)$$

$$L(\theta)=\prod_{i\in\{1,\ldots,N\},y^{(i)}=1}P(Y=1|X=x^{(i)};\theta)\cdot\prod_{i\in\{1,\ldots,N\},y^{(i)}=0}(1-P(Y=1|X=x^{(i)};\theta))$$

$$J(\theta)=-lnL(\theta)=-\sum_{i=1}^{N}y^{(i)}ln(P(Y=1|X=x^{(i)};\theta))+(1-y^{(i)})ln(1-P(Y=1|X=x^{(i)};\theta))$$

$$\nabla_\theta J(\theta)=\sum_i x^{(i)}(f(x^{(i)};\theta)-y^{(i)})$$

梯度法求函数取得最大值对应的参数值



似然函数

$$p\left(\boldsymbol{y}^{(i)}=1 \middle| \boldsymbol{X}^{(i)};\boldsymbol{w}\right)=h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)=\frac{1}{1+\ e^{-\boldsymbol{w}^T\boldsymbol{X}^{(i)}}}=\frac{1}{1+\ e^{-(\boldsymbol{w}_0\boldsymbol{X}_0^{(i)}+\boldsymbol{w}_1\boldsymbol{X}_1^{(i)}+\boldsymbol{w}_2\boldsymbol{X}_2^{(i)}+\cdots)}} \qquad \boldsymbol{X}_0^{(i)}\equiv 1$$

$$l(\boldsymbol{w})=\sum_{i=1}^{m}\boldsymbol{y}^{(i)}log h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)+\left(1-\boldsymbol{y}^{(i)}\right)log\left(1-h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right)$$

$$\frac{\partial l(\boldsymbol{w})}{\partial \boldsymbol{w}_j}=\sum_{i=1}^{m}\left(\boldsymbol{y}^{(i)}-h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right)\boldsymbol{X}_j^{(i)}$$

似然函数

$$p\left(\boldsymbol{y}^{(i)}=1\middle|\boldsymbol{X}^{(i)};\boldsymbol{w}\right)=h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)=\frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{X}^{(i)}}}=\frac{1}{1+e^{-(\boldsymbol{w}_0X_0^{(i)}+\boldsymbol{w}_1X_1^{(i)}+\boldsymbol{w}_2X_2^{(i)}+\cdots)}}$$

$$\boldsymbol{X}_0^{(i)}\equiv 1$$

$$\boldsymbol{L}(\boldsymbol{w})=p(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{w})=\prod_{i=1}^{m}p(\boldsymbol{y}^{(i)}|\boldsymbol{X}^{(i)};\boldsymbol{w})=\prod_{i=1}^{m}p(\boldsymbol{y}^{(i)}=1|\boldsymbol{X}^{(i)};\boldsymbol{w})^{\boldsymbol{y}^{(i)}}p(\boldsymbol{y}^{(i)}=0|\boldsymbol{X}^{(i)};\boldsymbol{w})^{1-\boldsymbol{y}^{(i)}}$$

$$=\prod_{i=1}^{m}p(\boldsymbol{y}^{(i)}=1|\boldsymbol{X}^{(i)};\boldsymbol{w})^{\boldsymbol{y}^{(i)}}\left(1-p(\boldsymbol{y}^{(i)}=1|\boldsymbol{X}^{(i)};\boldsymbol{w})\right)^{1-\boldsymbol{y}^{(i)}}=\prod_{i=1}^{m}\left(h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right)^{\boldsymbol{y}^{(i)}}\left(1-h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right)^{1-\boldsymbol{y}^{(i)}}$$

$$l(\boldsymbol{w})=log\boldsymbol{L}(\boldsymbol{w})=\sum_{i=1}^{m}\boldsymbol{y}^{(i)}logh_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)+\left(1-\boldsymbol{y}^{(i)}\right)log\left(1-h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right)$$

$$logAB=logA+logB;\quad logA^B=BlogA$$

似然函数，对权值求偏导

$$l(\boldsymbol{w}) = \sum_{i=1}^{m} \boldsymbol{y}^{(i)} log h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right) + \left(1 - \boldsymbol{y}^{(i)}\right) log \left(1 - h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right)$$

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

$$\frac{\partial h(z)}{\partial z} = h(z)\left(1 - h(z)\right)$$

$$h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right) \ = \frac{1}{1 + \ e^{-\boldsymbol{w}^T \boldsymbol{X}^{(i)}}}$$

$$\begin{aligned}
\frac{\partial l(\boldsymbol{w})}{\partial \boldsymbol{w}_j} &= \sum_{i=1}^{m} \left(\boldsymbol{y}^{(i)} \frac{1}{h_{\boldsymbol{w}}(\boldsymbol{X}^{(i)})} + \left(1 - \boldsymbol{y}^{(i)}\right) \frac{1}{1 - h_{\boldsymbol{w}}(\boldsymbol{X}^{(i)})} \right) \frac{\partial h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)}{\partial \boldsymbol{w}_j} \\
&= \sum_{i=1}^{m} \left(\boldsymbol{y}^{(i)} \frac{1}{h_{\boldsymbol{w}}(\boldsymbol{X}^{(i)})} + \left(1 - \boldsymbol{y}^{(i)}\right) \frac{1}{1 - h_{\boldsymbol{w}}(\boldsymbol{X}^{(i)})} \right) h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right) \left(1 - h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right) \frac{\partial \boldsymbol{w}^T \boldsymbol{X}^{(i)}}{\partial \boldsymbol{w}_j} \\
&= \sum_{i=1}^{m} \left(\boldsymbol{y}^{(i)} \left(1 - h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right)\right) + \left(1 - \boldsymbol{y}^{(i)}\right) h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right) \right) \boldsymbol{X}_j^{(i)} \\
&= \sum_{i=1}^{m} \left(\boldsymbol{y}^{(i)} - h_{\boldsymbol{w}}\left(\boldsymbol{X}^{(i)}\right) \right) \boldsymbol{X}_j^{(i)}
\end{aligned}$$

多元分类

$$h_{\mathbf{w}}(x) = \begin{bmatrix} P(y=1|x;\mathbf{w}) \\ P(y=2|x;\mathbf{w}) \\ \vdots \\ P(y=K|x;\mathbf{w}) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} \exp(w^{(j)\top} x)} \begin{bmatrix} \exp(w^{(1)\top} x) \\ \exp(w^{(2)\top} x) \\ \vdots \\ \exp(w^{(K)\top} x) \end{bmatrix}$$

$$\mathbf{w} = \begin{bmatrix} | & | & | & | \\ w^{(1)} & w^{(2)} & \cdots & w^{(K)} \\ | & | & | & | \end{bmatrix}$$

多元分类:似然函数及其导数

$$l(\mathbf{w}) = \sum_{i=1}^{n} \sum_{k=1}^{K} 1\left\{y^{(i)} = k\right\} \log \frac{\exp(w^{(k)\top} x^{(i)})}{\sum_{j=1}^{K} \exp(w^{(j)\top} x^{(i)})}$$

$$\nabla_{w^{(k)}} l(\mathbf{w}) = \sum_{i=1}^{n} \left[x^{(i)} \left(1\{y^{(i)} = k\} - P(y^{(i)} = k | x^{(i)}; w)\right)\right]$$

$$P(y^{(i)} = k | x^{(i)}; \mathbf{w}) = \frac{\exp(w^{(k)\top} x^{(i)})}{\sum_{j=1}^{K} \exp(w^{(j)\top} x^{(i)})}$$