回归算法——LR2

逻辑回归公式梳理

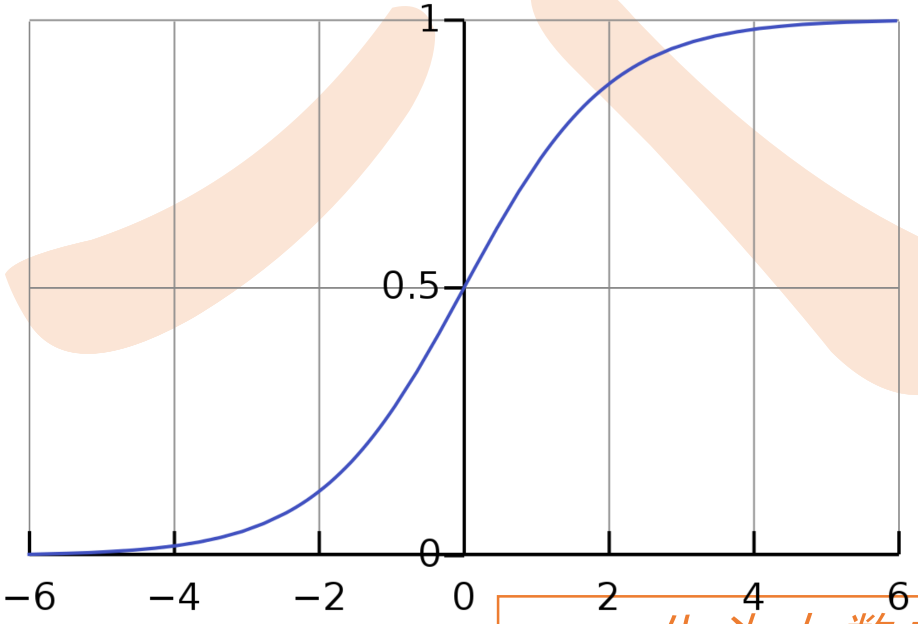
逻辑回归

定义一个条件概率：

$p(Y|X)$ 相当于用模型来捕获输入X和输出Y之间的关系

$$p(Y|X) = \mathrm{w}^T x + b$$ OK？

Logistic Function（sigmoid）：



$$y = \frac{1}{1 + e^{-x}}$$

$x: (-\infty, +\infty)$

$y: (0, 1)$

$$\sigma(x)$$

$$p(Y|X) = \sigma(\mathrm{w}^T x + b)$$

逻辑回归

$$p(Y|X) = \sigma(w^T x + b)$$

$$p(Y|X) = \frac{1}{1 + e^{-(w^T x + b)}}$$

$$w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \qquad w^T = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix}$$

$$b \in R$$

对于二分类问题：

$$p(y = 1|x, w) = \frac{1}{1 + e^{-(w^T x + b)}} \qquad p(y = 0|x, w) = \frac{e^{-(w^T x + b)}}{1 + e^{-(w^T x + b)}}$$

两个式子合并：

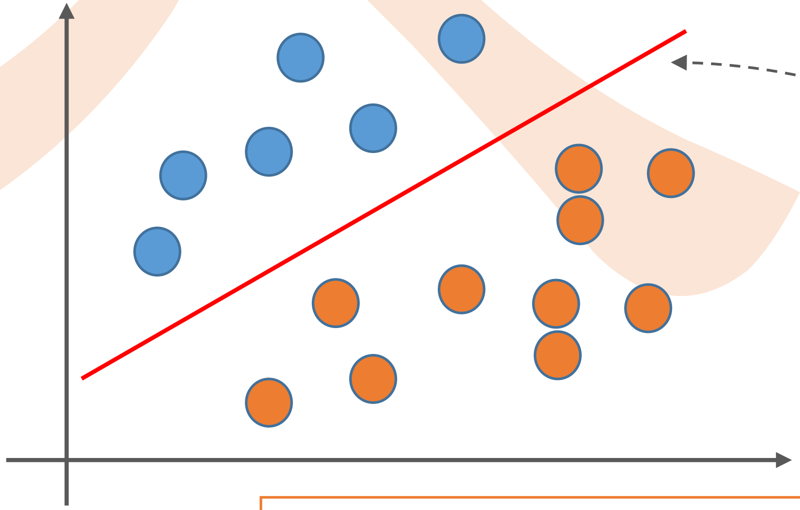
$$p(y|x, w) = p(y = 1|x, w)^y [1 - p(y = 1|x, w)]^{1-y}$$

逻辑回归

决策边界：

$$p(y=1|x,w)=\frac{1}{1+e^{-(w^Tx+b)}}$$

$$p(y=0|x,w)=\frac{e^{-(w^Tx+b)}}{1+e^{-(w^Tx+b)}}$$



$$\frac{p(y=1|x,w)}{p(y=0|x,w)}=1$$

$$\log\frac{p(y=1|x,w)}{p(y=0|x,w)}=\log 1$$

$$-(w^Tx+b)=0$$

$$w^Tx+b=0$$

逻辑回归

目标函数：

假设我们拥有数据集： $D = \{(x_i, y_i)\}_{i=1}^{n}$ $x_i \in R^d$ $y_i \in \{0,1\}$

前面归纳出： $p(y|x,w) = p(y=1|x,w)^y[1-p(y=1|x,w)]^{1-y}$

最大化目标函数：

$$\widehat{w}_{MLE}, \hat{b}_{MLE} = argmax_{w,b} \boxed{\prod_{i=1}^{n} p(y_i|x_i,w,b)}$$

最大化目标函数：

$$\hat{w}_{MLE}, \hat{b}_{MLE} = argmax_{w,b} \prod_{i=1}^{n} p(y_i|x_i,w,b)$$

$$= argmax_{w,b} \log(\textstyle\prod_{i=1}^{n} p(y_i|x_i,w,b))$$

$$= argmax_{w,b} \textstyle\sum_{i=1}^{n} \log(p(y_i|x_i,w,b))$$

- 引入log，避免类型精度越界

$$= argmin_{w,b} -\textstyle\sum_{i=1}^{n} \log(p(y_i|x_i,w,b))$$

- 找到概率最小值的w和b

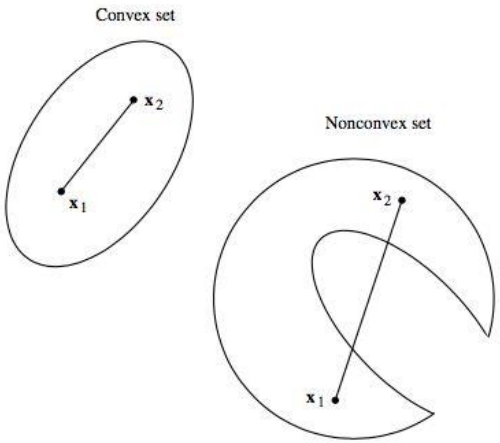
$$= argmin_{w,b} -\textstyle\sum_{i=1}^{n} \log(p(y_i=1|x_i,w,b)^{y_i}[1-p(y_i=1|x_i,w,b)]^{1-y_i})$$

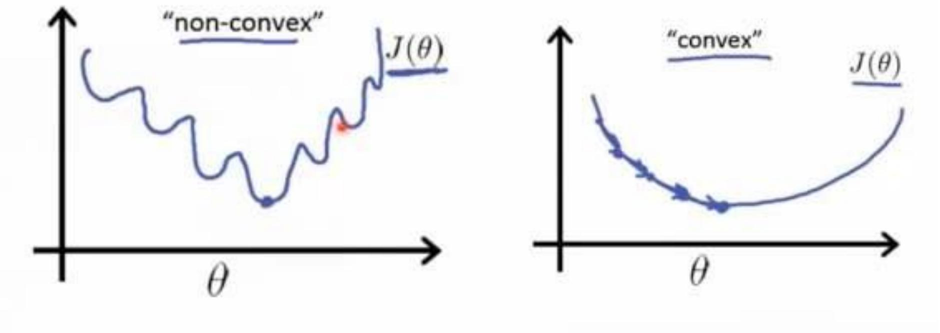
$$\boxed{\begin{aligned}\log(a^y b^x) &= \log(a^y) + \log(b^x)\\ &= yloga + xlogb\end{aligned}}$$

$$= argmin_{w,b} -\textstyle\sum_{i=1}^{n} y_i\log(p(y_i=1|x_i,w,b) + (1-y_i)\log[1-p(y_i=1|x_i,w,b)])$$

逻辑回归

凸函数：





优化算法：

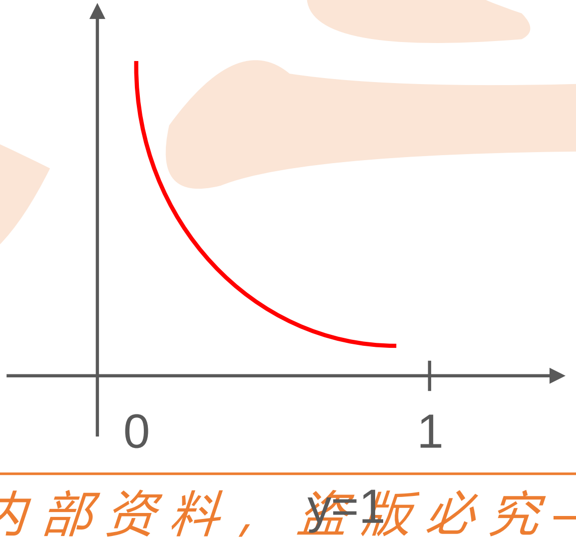
- GD（Gradient Descent）
- SGD（Stochastic Gradient Descent）

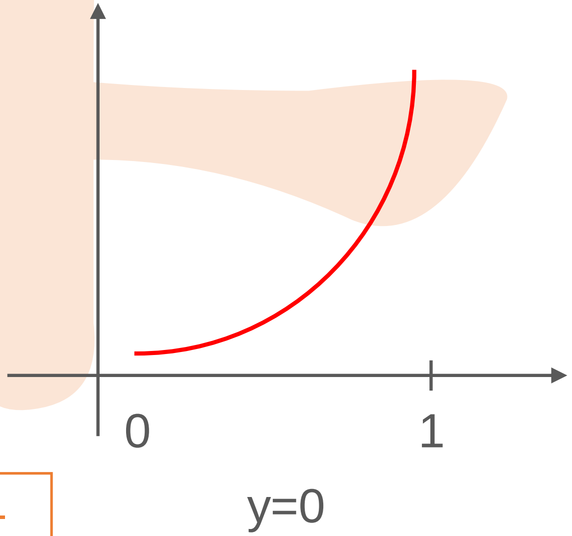
逻辑回归目标函数是凸函数

$$\hat{w}_{MLE}, \hat{b}_{MLE} = argmax_{w,b} \prod_{i=1}^{n} p(y_i|x_i, w, b)$$

$$= argmin_{w,b} - \sum_{i=1}^{n} y_i \log(p(y_i = 1|x_i, w, b) + (1 - y_i)\log[1 - p(y_i = 1|x_i, w, b)])$$

$$\mathrm{Cost}(\sigma(x)), y) = \begin{cases} -\log(\sigma(x)) & if\ y = 1 \\ -\log(1 - \sigma(x)) & if\ y = 0 \end{cases}$$



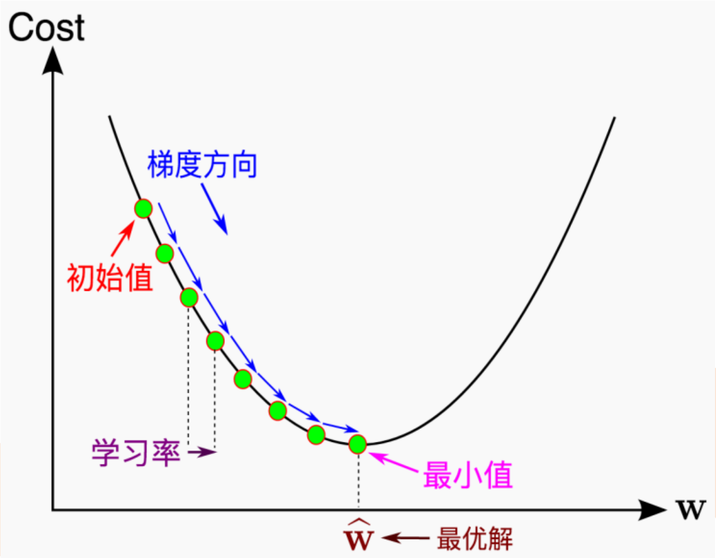


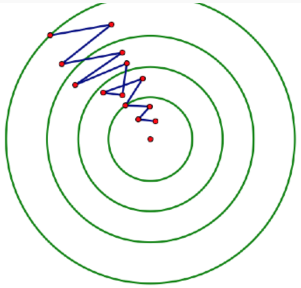
逻辑回归

例子：

求解f(x)值最小的对应参数w：

$$\text{初始化}w^1$$
$$for\ t = 1,2,\ldots$$
$$w^{t+1} = w^t - \eta \nabla f(w^t)$$





求解： $f(x) = 4w^2 + 5w + 1$ 最优解

$$w^* = -\frac{b}{2a} = -\frac{5}{8} = 0.625$$

初始化： $w^1 = 0 \qquad f'(x) = 8w + 5 \qquad \mu = 0.1$

$$w^2 = w^1 - 0.1(8*0+5) = -0.5$$
$$w^3 = w^2 - 0.1(8*(-0.5)+5) = -0.6$$
$$w^4 = w^3 - 0.1(8*(-0.6)+5) = -0.62$$
$$w^5 = w^4 - 0.1(8*(-0.62)+5) \approx -0.625$$

逻辑回归目标函数的梯度下降求解：

$$p(y=1|x,w)=\frac{1}{1+e^{-(w^Tx+b)}}=\sigma(w^Tx+\mathrm{b})$$

$$argmin_{w,b}-\sum_{i=1}^{n}y_i\log(p(y_i=1|x_i,w,b)+(1-y_i)\log[1-p(y_i=1|x_i,w,b)])$$

$$=argmin_{w,b}\underbrace{-\sum_{i=1}^{n}y_i\log(\sigma(w^Tx_i+b)\ +(1-y_i)\log[1-\sigma(w^Tx_i+b)])}_{L(w,b)}$$

$$\boxed{(logx)'=\frac{1}{x}}\qquad \boxed{\sigma(x)=\frac{1}{1+e^{-x}}}\qquad \boxed{\sigma'(x)=\sigma(x)[1-\sigma(x)]}$$

$$\partial\frac{L(w,b)}{w}=-\sum_{i=1}^{n}y_i\frac{\cancel{\sigma(w^Tx_i+b)}[1-\sigma(w^Tx_i+b)]}{\cancel{\sigma(w^Tx_i+b)}}x_i+(1-y_i)\frac{-\sigma(w^Tx_i+b)\cancel{[1-\sigma(w^Tx_i+b)]}}{\cancel{1-\sigma(w^Tx_i+b)}}x_i$$

$$=\sum_{i=1}^{n}[\sigma(w^Tx_i+b)-y_i]x_i$$

预测值　真实值

逻辑回归目标函数的梯度下降求解：

$$p(y=1|x,w)=\frac{1}{1+e^{-(w^Tx+b)}}=\sigma(w^Tx+\mathrm{b})$$

$$argmin_{w,b}-\sum_{i=1}^{n}y_i\log(p(y_i=1|x_i,w,b)+(1-y_i)\log[1-p(y_i=1|x_i,w,b)])$$

$$=argmin_{w,b}\underbrace{-\sum_{i=1}^{n}y_i\log(\sigma(w^Tx_i+b)\ +(1-y_i)\log[1-\sigma(w^Tx_i+b)])}_{L(w,b)}$$

$\boxed{(logx)'=\frac{1}{x}}$　　$\boxed{\sigma(x)=\frac{1}{1+e^{-x}}}$　　$\boxed{\sigma'(x)=\sigma(x)[1-\sigma(x)]}$

$$\partial\frac{L(w,b)}{b}=-\sum_{i=1}^{n}y_i\frac{\cancel{\sigma(w^Tx_i+b)}[1-\sigma(w^Tx_i+b)]}{\cancel{\sigma(w^Tx_i+b)}}+(1-y_i)\frac{-\sigma(w^Tx_i+b)\cancel{[1-\sigma(w^Tx_i+b)]}}{\cancel{1-\sigma(w^Tx_i+b)}}$$

$$=\sum_{i=1}^{n}[\sigma(w^Tx_i+b)-y_i]$$

逻辑回归目标函数的梯度下降求解：

- 梯度下降法：最小化F(w)
  - 1、设置初始w，计算F(w)
  - 2、计算梯度 $\nabla F(w)$
    - 下降方向：dir = $(-\nabla F(w))$
  - 3、 尝试梯度更新
    - $w^{new} = w + 步长 * dir$ 得到下降后的 w^{new}和 $F(w^{new})$
  - 4、如果 $F(w^{new}) - F(w)$较小，停止；

否则 $w = w^{new}; F(w) = F(w^{new})$ 跳到第2步

初始化w'，b'

$for\ t = 1,2,\dots$

$$w^{t+1} = w^t - \eta \sum_{i=1}^{n} [\sigma(w^T x_i + b) - y_i] x_i$$

$$b^{t+1} = b^t - \eta \sum_{i=1}^{n} [\sigma(w^T x_i + b) - y_i]$$

逻辑回归

随机梯度下降求解：

$初始化w'，\ b'$

$for\ t = 1,2, \ldots$

$$w^{t+1} = w^{t} - \eta \cancel{\sum_{i=1}^{n}} [\sigma(w^{T}x_i + b) - y_i]x_i$$

$$b^{t+1} = b^{t} - \eta \cancel{\sum_{i=1}^{n}} [\sigma(w^{T}x_i + b) - y_i]$$

- 每次从训练样本中抽取一个样本进行更新，每次都不用遍历所有数据集，迭代速度快，需要迭代更多次数
- 每次选取方向不一定是最优

批量梯度下降和随机梯度下降，折中的方法：Mini-batch Gradient Descent

逻辑回归

进一步分析：

现象： 线性可分，逻辑回归参数趋近于无穷大！

$$p(y=1|x,w)=\frac{1}{1+e^{-(w^Tx+b)}}$$

$$p(y=0|x,w)=\frac{e^{-(w^Tx+b)}}{1+e^{-(w^Tx+b)}}$$

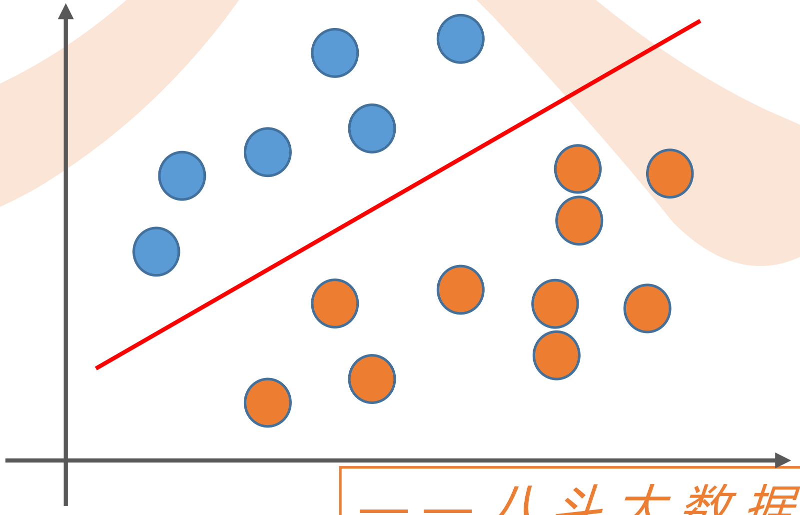
如果w非常大 →

$$p(y=1|x,w)\approx 1$$

$$p(y=0|x,w)\approx 0$$

非常理想！

过拟合现象



正则化：

目标：避免w过大

$||w||_2^2 = w_1^2 + w_2^2 + \ldots + w_d^2$

L2-norm: $\hat{w}_{MLE}, \hat{b}_{MLE} = argmin_{w,b}(-\prod_{i=1}^{n} p(y_i|x_i, w, b) + \lambda||w||_2^2)$

- λ：超参数，注意力权衡作用

逻辑回归

正则化：

目标： 避免w过大

目标函数：

$$\hat{w}_{MLE},\hat{b}_{MLE} = argmin_{w,b}(\underbrace{-\prod_{i=1}^{n} p(y_i|x_i,w,b)}_{L(w,b)} + \underbrace{\lambda||w||_2^2}_{R(w)})$$

$$\partial\frac{L(w,b)}{w} = \sum_{i=1}^{n}[\sigma(w^T x_i + b) - y_i]x_i$$

$$\boxed{\frac{\partial||w||_2^2}{\partial w} = 2w}$$

$$\partial\frac{L(w,b)+R(w)}{w} = \sum_{i=1}^{n}[\sigma(w^T x_i + b) - y_i]x_i + 2\lambda w$$

Q&A

@八斗数据