一、 逻辑回归算法简介

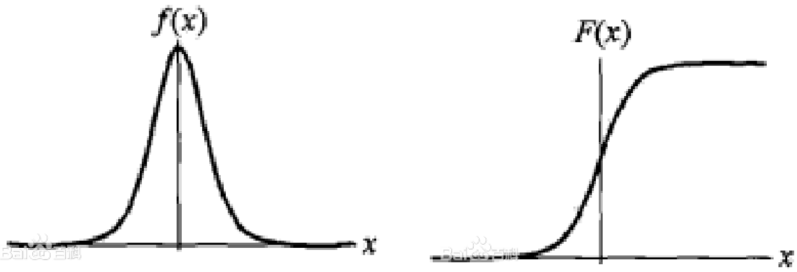
逻辑回归是一个经典的分类方法，逻辑回归模型是一个对数线性模型。为了研究逻辑回归算法，我们首先来学习什么是逻辑分布。

设 X 为连续随机变量，X 符合以下分布函数和密度函数：

$$F(x) = P(X \le x) = \frac{1}{1 + e^{-(x-\mu)/\gamma}}$$

$$f(x) = F'(x) = \frac{e^{-(x-\mu)/\gamma}}{\gamma(1 + e^{-(x-\mu)/\gamma})^2}$$

其中，μ为位置参数，γ>0 为形状参数。这两个函数的图像如下图所示：



根据分布函数图像我们很容易看出，曲线在中心附近增长速度比较快，两端增长速度比较慢。形状参数γ的值越小，曲线在中心附近增长的越快。逻辑分布就是逻辑回归算法的基础。

对于逻辑回归，主要的思路就是面对一个分类问题时，先建立代价函数，然后通过优化方法迭代求出最优的参数模型，再测试并验证这个模型的好坏。

二、 逻辑回归算法流程

1. 构造预测函数

逻辑分布的函数实际上就是一个 sigmoid 函数 ,我们记作$g(x) = \frac{1}{1+e^{-z}}$，当它是一个线性边界的时候，边界形式可以记作

$$z = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n = \sum_{i=0}^{n} \theta_i x_i$$

我们需要的数据就是最佳参数

$$\theta = [\theta_0, \theta_1, \theta_2, \ldots, \theta_n]^T$$

构造预测函数为

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

函数 h(x)的值有特殊的含义，他表示结果取 1 的概率。因此对于输入 x 分类结果为类别 1 和类别 0 的概率分别为：$P(y=1|x_\theta) = h_\theta(x)$，$P(y=0|x_\theta) = 1 - h_\theta(x)$

2. 构造损失函数

在构造损失函数时，我们利用极大似然估计的方法。通俗一点的说，极大似然估计，就是利用已知的样本结果，反推出最有可能，也就是最大概率导致这样结果的模型参数值。

设$P(Y=1|x) = \pi(x)$, $P(Y=0|x) = 1 - \pi(x)$,则似然函数可表示为：

$$\prod_{i=1}^{N} [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i}$$

它的对数似然函数为：

$$\begin{aligned} L(\omega) &= \sum_{i=1}^{N} [y_i \, log \, \pi(x_i) + (1 - y_i) \, log(1 - \pi(x_i))] \\ &= \sum_{i=1}^{N} [y_i log \frac{\pi(x_i)}{1 - \pi(x_i)} + log \, (1 - \pi(x_i))] \\ &= \sum_{i=1}^{N} [y_i(\omega \cdot x_i) - log(1 + exp(\omega \cdot x_i))] \end{aligned}$$

对$L(\omega)$求极大值，就可以得到ω的估计值。这样问题就被转化为以对数似然函数为目标函数的最优化问题。假设求出的极大似然估计值是$\widehat{\omega}$，那么学习到的逻辑回归模型为

$$P(Y=1|x)=\frac{exp\,(\widehat{\omega}\cdot x)}{1+exp\,(\widehat{\omega}\cdot x)}$$
$$P(Y=0|x)=\frac{1}{1+exp\,(\widehat{\omega}\cdot x)}$$

三、 多项逻辑回归

前面的逻辑回归是二项分类模型，用于二分类的问题。但这种思想也适用于多分类模型，称为多项逻辑回归模型。关于二项逻辑回归的参数估计法，也可以推广到多项逻辑回归模型中使用。

四、 逻辑回归算法优缺点总结

优点

- 计算代价低，易于理解和实现；
- 速度快，适合解决二分类问题；
- 能够容易地更新模型并吸收新的数据；

缺点

- 对数据和场景的适应性有限；
- 容易欠拟合，分类精读相对不高；