

逻辑回归模型与线性回归模型

逻辑回归模型与线性回归模型都是广义上的线性模型

相同点：逻辑回归（Logistic Regression）去掉最后一层Sigmoid激活函数后，即为线性回归（Linear Regression）的预测值。

不同点：

- 预测值y分布不同：逻辑回归会假设因变量y服从伯努利的二项分布，线性回归模型假设因变量y服从高斯分布；
- 损失函数不同：逻辑回归的loss函数是二值或多值的交叉熵损失函数，而线性回归模型的loss函数是均方差损失函数；
- 优化方法不同：逻辑回归的优化方法为最大似然概率，线性回归模型的优化方法为最小二乘法。
- 模型目标不同：逻辑回归的目标是估计某个类别事件发生的可能性，常用于解决二分类问题，即条件概率 $P(y=1|x)>0.5$? 这里阈值常选用0.5，如果想提高正样本的准确率，可以适当提高这个阈值，反之如果想提高正样本的召回率，可以适当降低这个阈值。线性回归模型的目标利用多元维度特征是拟合出一条线段来代表样本数据分布情况，常用于回归预测未来某一时刻的数值。

逻辑回归模型

逻辑回归模型分成概率值建模和参数估计两部分，概率值建模是指建模函数及其函数概率分布，而参数估计为模型参数优化理论。

1.逻辑回归建模

逻辑回归模型分为二值和多值的概率预测，为了方便说明其模型原理，这里只讲二值的条件概率预估模型，条件概率分布 $P(Y|X)$ 表示预测的概率，其中Y的取值为1或0。

二值的逻辑回归模型满足如下条件概率分布：

$$P(Y = 1|X) = \frac{\exp^{wx+b}}{1 + \exp^{wx+b}}; P(Y = 0|X) = 1 - P(Y = 1|X) = \frac{1}{1 + \exp^{wx+b}}$$

其中w代表x变量权重，b代表整体偏值。（如果x有多个则扩展为 $w_1*x_1+w_2*x_2+...$ ）。通过比较两个条件概率的大小，将实例x分到概率值较大的那一类。

$$P(Y = 1|X) = \frac{\exp^{wx+b}}{1 + \exp^{wx+b}};$$

条件概率 $P(Y=1|X)$ ，这时，线性函数 $wx+b$ 的值越接近正无穷，概率值就越接近1；线性函数的值越接近负无穷，概率值就越接近0，这样的模型就是Logistic回归模型。

2.逻辑回归参数估计

Logistic回归模型是一种概率模型，通常采用极大似然估计法来求解模型参数，对于给定的训练集 $T=\{(x_1, y_1); (x_2, y_2); (x_3, y_3); \dots (x_n, y_n)\}$ ，其中 x_i 属于任意实数， y_i 取值为0或1。

假设 $P(Y=1|x) = \pi(x)$, $P(Y=0|x) = 1-\pi(x)$ ，则有

$$P(Y|x) = \pi(x)^y(1 - \pi(x))^{1-y}$$

由于随机变量Y为离散型随机变量，其联合概率分布函数即为似然函数

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

对数似然函数为

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

对于L(w)求极大值，即导数等于0，分别得到w的估计和b的估计。利用随机梯度下降法或拟牛顿方法，求解极大似然估计值为w'和b'，对应的模型为：

$$P(Y = 1|X) = \frac{\exp^{w'x+b'}}{1 + \exp^{w'x+b'}}; P(Y = 0|X) = 1 - P(Y = 1|X) = \frac{1}{1 + \exp^{w'x+b'}}$$

线性回归模型

线性回归模型一般用于预测未来某一时刻的数值，是一个基本回归问题。给定特定的m个训练样本D={x0,x1,x2,...xd;ym}线性回归试图学得一个通过属性的线性组合来进行预测的函数，即

$$f(x) = w_1x_1 + w_2x_2 + \dots + w_dx_d + b$$

上述式子可简化为

$$f(x_i) = wx_i + b, \text{使得 } f(x_i) \approx y_i$$

怎样衡量f(xi)到yi的距离呢，可以利用均方误差MSE，即欧式距离

$$\begin{aligned} (w^*, b^*) &= \operatorname{argmin}_{w,b} \sum_{i=1}^m (f(x_i) - y_i)^2 \\ &= \operatorname{argmin}_{w,b} \sum_{i=1}^m (y_i - wx_i - b)^2 \end{aligned}$$

基于均方误差最小化来进行模型求解的方法称为“最小二乘法”，在线性回归中，最小二乘法就是试图找到一条直线，使所有样本到直线上的[欧式距离](#)之和最小，求解w和b使上面公式最小化的过程，称为线性回归模型的最小二乘“参数估计”，分别对w和b求导可得

$$\begin{aligned} \frac{\partial E(w, b)}{\partial w} &= 2(w \sum_{i=1}^m x_i^2 - \sum_{i=1}^m (y_i - b)x_i) \\ \frac{\partial E(w, b)}{\partial b} &= 2(mb - \sum_{i=1}^m (y_i - wx_i)) \end{aligned}$$

令上面两条公式为零可得到w和b最优解的闭式解

$$\begin{aligned} w &= \frac{\sum_{i=1}^m y_i(x_i - \bar{x})}{\sum_{i=1}^m (x_i^2 - \frac{1}{m}(\sum_{i=1}^m x_i)^2)} \\ b &= \frac{1}{m} \sum_{i=1}^m (y_i - wx_i) \end{aligned}$$

仔细观察上述等式，事实上 w 是在逐步渐进样本均值和方差的过程，而 b 则是在校准偏移量过程。实际上多层多元的线性回归加上非线性激活函数就是如今的神经网络拟合数值的过程，不难理解神经网络拟合的效果要比多元线性回归好很多。