

机器学习常见知识汇总之线性模型

田晓彬

xiaobin9652@163.com

线性回归

线性回归试图学习得到一个线性模型以尽可能准确地预测输入值的输出标记。为了简化描述，我们讨论输入值只有一个属性的二分类情况， $D = \{(x_i, y_i) | i \in 1, 2, \dots, n\}$ 。

我们很容易可以得到 y 关于 x 的线性模型 $y_i = wx_i + b$ 。但在实际应用中我们不一定能找到完全合适的参数 w 、 b 来使等式完全成立，所以我们将线性回归的学习目标定义为

$$f(x_i) = wx_i + b \quad \text{s.t. } f(x_i) \approx y_i.$$

接下来需要确定参数 w 、 b ，在此引入均方误差来度量 $f(x_i)$ 和 y_i 之间的差距，并最小化均方误差来确定最优的参数

$$\arg \min_{w, b} \sum_{i=1}^n (f(x_i) - y_i)^2 = \arg \min_{w, b} \sum_{i=1}^n (y_i - wx_i - b)^2.$$

有均方误差的公式可以看出，均方误差具有很好的几何意义，即对应了几何上常用的欧几里得距离。在线性回归中引入均方误差，即试图找到最优参数对应的一条直线，使得所有样本点到该直线的欧式距离之和最小。线性回归的优化函数可以通过最小二乘法进行求解，具体的求解步骤在此不再叙述。

逻辑回归

1. 逻辑回归不是回归问题

输入变量和输出变量均为连续变量的预测问题称之为回归问题；输出变量为有限个离散变量的预测问题称之为分类问题。

虽然逻辑回归叫做回归，但是它并不是一个回归问题。在逻辑回归中，因变量取值是一个二元分布，模型学习得到的是 $E[y|x; \theta]$ ，即给定自变量和超参数以后，得到因变量的期望。

2.逻辑回归的详细解释

在线性回归中，我们通过求解 $z = \theta x + \mu$ 来近似 z ，其中 μ 是误差项。若是希望用回归来处理分类任务，我们需要将线性回归系统的输出 z 转换成0/1值。首先想到的就是单位跃阶函数：

$$y = \begin{cases} 0, & z < 0; \\ 0.5, & z = 0; \\ 1, & z > 0, \end{cases}$$

其中0叫做阈值（可以根据具体情况换成其它值），若预测值大于阈值则为正例，若预测值小于阈值则为负例，若预测值等于阈值则可以任意判别。

但是很显然，单位跃阶函数是不可微的，我们希望找到一个与单位跃阶函数相似又可微的函数。*Sigmoid*函数定义如下：

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

其中 $z = w^T x + b$ 。由于*sigmoid*函数的特性，它输出的结果不再是预测结果，而是预测结果为正例的概率，预测结果为负例的概率为：

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

两者的比值

$$\frac{y}{1 - y} = \frac{1}{e^{-z}}$$

就是 x 作为正例的可能性，成为“几率”。对几率取对数则得到对数几率：

$$\ln \frac{y}{1 - y} = z = w^T x + b$$

若将 $f(z)$ 视为后验概率 $p(y = 1|x)$ ，

$$\ln \frac{p(y = 1 | x)}{p(y = 0 | x)} = w^T x + b$$

显然有

$$p(y = 1 | x) = \frac{e^{w^T x}}{1 + e^{w^T x}}$$

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

由概率论可知，

$$p(y_i | x_i; w, b) = (p(y = 1 | x; w, b))^{y_i} \times (1 - p(y = 0 | x; w, b))^{1-y_i}$$

根据最大似然估计可知，对于很多个样本 x_i ，若希望求得最合适的参数 w, b 使正确预测的概率最大，则求得合适的参数使得所有样本正确预测概率的乘积最大，即：

$$l(w, b) = \prod_{i=1}^N p(y_i | x_i; w, b)$$

等式两边同时取对数似然

$$L(w, b) = \ln l(w, b) = \sum_{i=1}^N (y_i \ln p_1(x_i; w, b) + (1 - y_i) \ln p_0(x_i; w, b))$$

化简可得

$$L(w, b) = \sum_{i=1}^N (y_i(w^T x + b) - \ln(1 + e^{w^T x + b}))$$

由于我们习惯于最小化一个函数，所以 $L(w, b)$ 划为

$$\min_{w, b} L(w, b) = \sum_{i=1}^N (-y_i(w^T x + b) + \ln(1 + e^{w^T x + b}))$$

这个函数就是逻辑回归的损失函数，被叫做交叉熵损失函数。

3.逻辑回归损失函数的求解

求解该损失函数需要使用梯度下降方法，在此不做详细介绍。

4.逻辑回归适用性

1. 用于概率预测：逻辑回归的输出为正例的可能性，即概率；
2. 用于分类问题：逻辑回归的输出虽然为概率，但是我们可以通过添加阈值来对逻辑回归的输出进行分类。不同的分类问题可能对应不同的阈值，需要根据具体情况来进行具体确定；
3. 仅能用于线性问题：只有当目标和特征是线性关系时，才能使用逻辑回归；

4. 特征不需要满足独立条件假设：各个特征之间不需要满足条件独立假设，但是各个特征的贡献需要独立计算。

5. 逻辑回归和线性回归的异同。

逻辑回归是分类问题，线性回归是回归问题，这是两者最本质的区别；逻辑回归中的因变量是离散的，而线性回归中的因变量是连续的；在自变量 x 和超参数 w, b 确定的情况下，逻辑回归可以看做广义线性模型在因变量 y 服从二元分布的一个特殊情况，线性回归的求解使用最小二乘法，可以认为是因变量 y 服从正态分布。

两者也有相同的地方，首先二者都是在使用极大似然估计来对训练样本进行建模。线性回归使用最小二乘法进行求解，实际上就是在自变量 x 和超参数 w, b 确定、因变量 y 服从正态分布的情况下，使用极大似然估计的一个化简；而逻辑回归中通过对似然函数

$$l(w, b) = \prod_{i=1}^N p(y_i | x_i; w, b) = \prod_{i=1}^N (p(x_i)^{y_i} (1 - p(x_i))^{(1-y_i)})$$

的学习，得到最优超参数 w, b 。另外，二者在求解超参数的过程中，都可以使用梯度下降算法。