

逻辑回归二分类总结(不断迭代，直到彻底掌握)



凡人求索

数据挖掘农民工

关注他

9 人赞同了该文章

——背景——

近来，想把学习过的机器学习算法做一个总结，于是打算结合网上的面试经验和工作的需求写下一个机器学习系列文章。这里是更多是参考网上以及书本中的知识，目的系统的整理这些知识，并且只要是参考的文章都会给相应的链接，大家也可以根据需要进行学习。

——正文——

用一句话来介绍逻辑回归就是这样的，假设数据服从**伯努利分布[1]**，通过**极大化似然估计[2]**的方法，运用**梯度下降法[3]**来求解参数，来达到将数据二分类的目的。

此篇文章会由为什么不用线性回归模型做分类来引入逻辑回归,从逻辑回归的假设，到逻辑回归的损失函数，到逻辑回归的求解依次来展开，结尾会提出逻辑回归一些细节的问题并给出相应文章的参考链接。

[1] 为什么不用线性回归模型做分类？

为什么不用线性回归模型，而是使用逻辑斯蒂回归模型，是因为线性回归用于二分类时会出现这样的问题：如判定一封邮件是否为垃圾邮件，我们需要通过分类器预测分类结果：

▲ 赞同 9 ▼

$$p = f_{\theta}(x) = \theta^T x$$

线性回归

这里存在的问题是：

(1)等式两边的取值范围不同，右边是负无穷到正无穷，左边是[0,1],显然这样的分类模型存在着问题

(2)实际中的很多问题，都是当x很小或是很大时，对于因变量P的影响很小，当x达到中间某个阈值时，影响很大。我们拿买车的例子来看，当一个人从没有钱到有钱，他买车的概率就会上升，等有钱到一定程度之后，随着挣钱的增加，他买车的概率应该就不会再增加了。即实际中很多问题，概率P与自变量并不是直线关系。

所以，逻辑回归引入logistic函数，使其成为非线性关系，并将分类器输出界定在 [0 , 1] 之间，其一般形式可表示为

$$f_{\theta}(x) = g(\theta^T x)$$

那么问题来了，我们到底需要选择什么样形式的logistic函数g(z)？

首先线性回归不能保证预测值的范围位于 [0 , 1] 之间。所以，针对该问题最简单的方法是移除因变量值域的限制，这样，我们就需要对概率形式做某种变换。首先，选用优势比Odds代替概率，优势比是事件发生概率和不发生概率之间的比值，记为

$$odds = \frac{p}{1-p}$$

通过该变换，我们可以将 [0 , 1] 之间的任意数映射到[0,∞]之间的任意实数。但是，线性回归的输出还可以是负数，我们还需要另一步变换将[0,∞]的实数域映射到这个实数域R空间；

然后，在众多非线性函数中，log函数的值域为整个实数域且单调，因此，我们可以计算优势比的对数。

$$\eta = \log(odds) = \log \frac{p}{1-p} = \text{logit}(p)$$

经过以上两步，我们可以去除分类问题对因变量值域的限制，如果概率等于0，那么优势比则为0，logit值为 $-\infty$ ；相反，如果概率等于1，优势比为 ∞ ，logit值为 ∞ 。因此，logi

▲ 赞同 9 ▼

综上所述,我们解决了分类问题对分类器预测的因变量值域的限制,我们就可以采用线性回归对一一映射后的概率值进行线性拟合,即

$$\log it(p) = \log\left(\frac{p}{1-p}\right) = \eta = f_{\theta}(x) = \theta^T x$$

因为logit变换是一一映射,所以存在logit的反变换(antilogit),我们可以求得p值的解析表达式:

$$p = \text{anti log it}(x) = \log it^{-1} = \frac{e^{\eta}}{1 + e^{\eta}} = \frac{e^{\theta^T x}}{1 + e^{\theta^T x}}$$

$$1-p = 1 - \text{anti log it}(x) = 1 - \log it^{-1} = 1 - \frac{e^{\eta}}{1 + e^{\eta}} = 1 - \frac{e^{\theta^T x}}{1 + e^{\theta^T x}} = \frac{1}{1 + e^{\theta^T x}}$$

上式即为logistic其中采用的logit的反变换一般记为sigmoid函数:

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

[补充]: 此处涉及到两个假设,第一个基本假设是:假设数据服从伯努利分布,这里二分类器的假设是,给定 x 后, y 是符合伯努利分布的。对于不同的 x 值, y 就符合不同均值的伯努利分布。第二个假设是:样本为正(即是1)的概率是

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

[2] 逻辑回归的损失函数

通过以上得知这两个公式,且其中 θ 是未知参数

$$P(y = 1|x, \theta) = h_{\theta}(x)$$

$$P(y = 0|x, \theta) = 1 - h_{\theta}(x)$$

注意:这里的 $h(x)$ 即是上文的 $g(x)$

知乎

(Maximum Likelihood Method) 是建立在各样本间相互独立且样本满足随机抽样 (可代表总体分布) 下的估计方法，它的核心思想是如果现有样本可以代表总体，那么最大似然估计就是找到一组参数使得出现现有样本的可能性最大，即从统计学角度需要使得所有观测样本的联合概率最大化，又因为样本间是相互独立的，所以所有观测样本的联合概率可以写成各样本出现概率的连乘积，如下式：

$$L(\theta) = \prod_{i=1}^m (h_{\theta}(x^{(i)})^{y^{(i)}} (1 - h_{\theta}(x^{(i)}))^{1-y^{(i)}})$$

我们知道线性回归可以采用梯度下降算法求解,是因为线性回归的代价函数 (均方误差函数) 是凸函数，为碗状，而凸函数具有良好的性质 (对于凸函数来说局部最小值点即为全局最小值点) 使得我们一般会将非凸函数转换为凸函数进行求解。因此，最大似然估计采用自然对数变换，将似然函数转换为对数似然函数，且相对于求解对数似然函数的最大值，我们当然可以将该目标转换为对偶问题，即求解代价函数 $J(\theta) = -\log(l(\theta))$ 的最小值。因此，我们定义**logistic回归的代价函数[4]**为：

$$J(\theta) = -\ln L(\theta) = -\sum_{i=1}^m (y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})))$$

损失函数用矩阵法表达更加简洁：

$$J(\theta) = -Y \bullet \log h_{\theta}(X) - (E - Y) \bullet \log(E - h_{\theta}(X))$$

其中E为单位向量,•为内积。

对目标函数求解，**求解目标函数的一阶偏导数获得梯度[5]**，为：

这一步我们用到了矩阵求导的链式法则，和下面三个矩阵求导公式：

$$\frac{\partial}{\partial X} \log X = 1/X$$

$$\frac{\partial}{\partial z} g(z) = g(z)(1 - g(z)) \quad (g(z) \text{ 为 sigmoid 函数})$$

$$\frac{\partial}{\partial \theta} X\theta = X^T$$

对于刚才的求导公式我们进行化简可得：

$$\frac{\partial}{\partial \theta} J(\theta) = X^T(h_{\theta}(X) - Y)$$

从而在梯度下降法中每一步向里 θ 的迭代公式如下：

$$\theta = \theta - \alpha X^T(h_{\theta}(X) - Y)$$

其中， α 为梯度下降法的步长。

[3]逻辑回归如何分类

逻辑回归是如何应用到分类上去呢？。我们知道预测出的结果是样本为正的的概率，这是一个连续值，我们要做的是划定一个阈值，P值大于这个阈值的是一类，P值小于这个阈值的是另外一类。阈值具体如何调整根据实际情况选择。一般会选择0.5做为阈值来划分。

——逻辑回归的一些常见问题——

[1].你在使用逻辑回归的时候有哪些感受。觉得它有哪些优缺点？

逻辑回归的优缺点

[2.]逻辑回归更深入的思考(这是在看spass modler有关资料所得)

a. 什么时候我们能够使用逻辑回归呢？

$$\log it(p) = \log\left(\frac{p}{1-p}\right) = \eta = f_{\theta}(x) = \theta^T x$$

其实我们得出这个公式的时候，我们做了变量全体是和logit(P)是有线性关系的假设。所以我们要做回归方程显著性检验，目的就是检验解释变量全体是和logit(P)的线性关系是否显著，是否可以用线性模型拟合。其原假设H(0)是：各回归系数同时为0，解释变量全体与logit(P)的线性关系不显著。如果是不显著的话，我们可以尝试用特征离散化解决非线性特征问题[6]。

b.二项逻辑回归方程系数的意义[7]

c.回归系数也是要做显著性检验的，目的是逐个检验方程中各个解释变量是否与logit(P)有显著的线性关系，对logit(p)是否有重要贡献。

d.模型效果的检验[8]

(此篇文章会不断的更新)

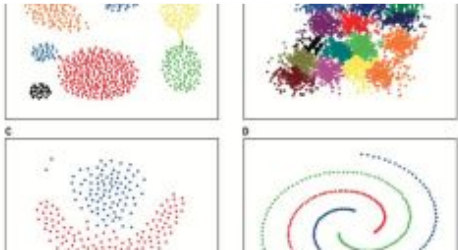
参考

- 1:[通俗的语言解释一下伯努利分布和二项分布的区别\(知乎\)](#)
- 2:[一文搞懂极大似然估计](#)
- 3:[梯度下降小结](#)
- 4:[逻辑回归原理小结](#)
- 5:[矩阵求导](#)
- 6:[特征离散化解决非线性特征问题](#)
- 7:[深入解读Logistic回归结果（一）：回归系数，OR](#)
- 8:[怎样用SPSS做二项Logistic回归分析？结果如何解释？](#)

编辑于 2018-03-04

[数据挖掘](#) [数据分析](#)

推荐阅读



数据科学家需要了解的5种聚类算法

论智

▲ 赞同 9 ▼

1 条评论

切换为时间排序

写下你的评论...



徐松

9 个月前

对逻辑回归的公式怎么来的讲的很清楚，感谢~

1