

逻辑回归模型(Logistic Regression, LR)基础

逻辑回归(Logistic Regression, LR)模型其实仅在线性回归的基础上,套用了一个逻辑函数,但也就由于这个逻辑函数,使得逻辑回归模型成为了机器学习领域一颗耀眼的明星,更是计算广告学的核心。本文主要详述逻辑回归模型的基础,至于逻辑回归模型的优化、逻辑回归与计算广告学等,请关注后续文章。

1 逻辑回归模型

回归是一种极易理解的模型,就相当于 $y=f(x)$,表明自变量 x 与因变量 y 的关系。最常见问题有如医生治病时的望、闻、问、切,之后判定病人是否生病或生了什么病,其中的是获取自变量 x ,即特征数据,判断是否生病就相当于获取因变量 y ,即预测分类。

最简单的回归是线性回归,在此借用Andrew NG的讲义,有如图1.a所示, X 为数据点——肿瘤的大小, Y 为观测值——是否是恶性肿瘤。通过构建线性回归模型,如 $h_{\theta}(x)$ 所示:回归模型后,即可以根据肿瘤大小,预测是否为恶性肿瘤 $h_{\theta}(x) \geq 0.5$ 为恶性, $h_{\theta}(x) < 0.5$ 为良性。

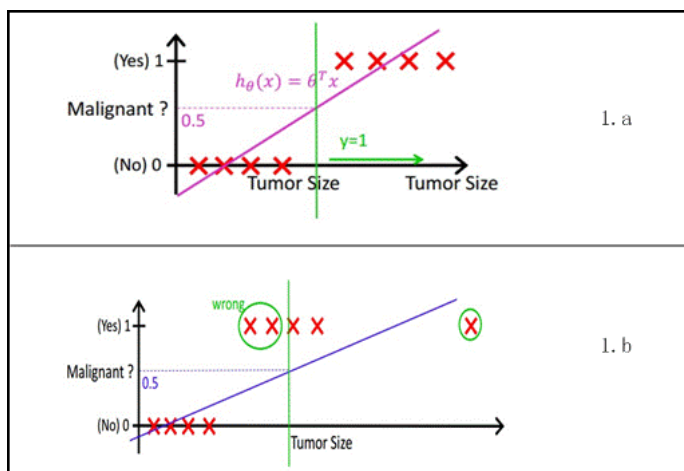


图1 线性回归示例

然而线性回归的鲁棒性很差,例如在图1.b的数据集上建立回归,因最右边噪点的存在,使回归模型在训练集上表现都很差。这主要是由于线性回归在整个实数域内敏感度一致围,需要在 $[0,1]$ 。逻辑回归就是一种减小预测范围,将预测值限定为 $[0,1]$ 间的一种回归模型,其回归方程与回归曲线如图2所示。逻辑曲线在 $z=0$ 时,十分敏感,在 $z > 0$ 或 $z < 0$ 敏感,将预测值限定为 $(0,1)$ 。

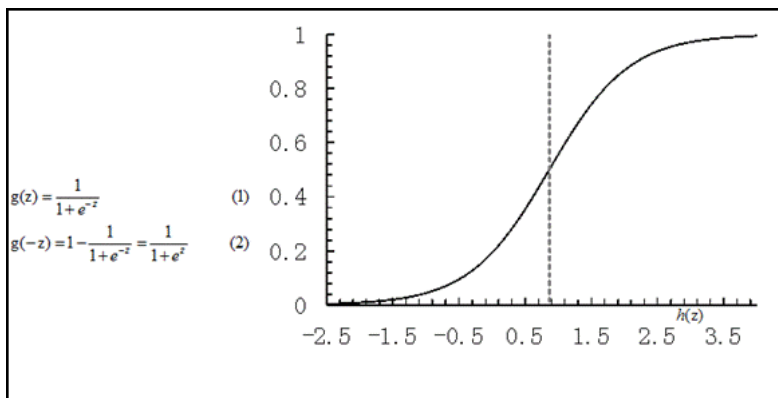


图2 逻辑方程与逻辑曲线

逻辑回归其实仅为在线性回归的基础上,套用了一个逻辑函数,但也就由于这个逻辑函数,逻辑回归成为了机器学习领域一颗耀眼的明星,更是计算广告学的核心。对于多元回归如下公式拟合分类,其中公式(4)的变换,将在逻辑回归模型参数估计时,化简公式带来很多益处, $y = \{0,1\}$ 为分类结果。

$$\begin{cases} p(y=1|x, \theta) = \frac{1}{1+e^{-\theta^T x}} & (3) \\ p(y=0|x, \theta) = \frac{1}{1+e^{-\theta^T x}} = 1 - p(y=1|x, \theta) = p(y=1|x, -\theta) & (4) \end{cases}$$

$$\text{令: } h_{\theta}(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}; \quad g(z) = \frac{1}{1+e^{-z}}$$

对于训练数据集,特征数据 $x = \{x_1, x_2, \dots, x_m\}$ 和对应的分类数据 $y = \{y_1, y_2, \dots, y_m\}$ 。构建逻辑回归模型 $f(\theta)$,最典型的构建方法便是应用极大似然估计。首先,对于单后验概率为:

$$p(y|x, \theta) = (h_{\theta}(x))^y (1 - h_{\theta}(x))^{1-y}$$

其中 $y = 1$ (或0)

那么,极大似然函数为:

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

log似然是：

$$l(\theta) = \log(L(\theta | x, y)) \\ = \sum_{i=1}^m y^{(i)} \log(h(x^{(i)})) + (1 - y^{(i)}) \log(1 - h(x^{(i)}))$$

2 梯度下降

由第1节可知，求逻辑回归模型f(θ)，等价于：

$$\theta^* = \arg \min_{\theta} (l(\theta))$$

采用梯度下降法：

$$\frac{\partial}{\partial \theta_j} (l(\theta)) = \frac{\partial}{\partial \theta_j} \left(\sum_{i=1}^m y^{(i)} \log(h(x^{(i)})) + (1 - y^{(i)}) \log(1 - h(x^{(i)})) \right) \\ = \left(\frac{y^{(i)}}{h(x^{(i)})} - (1 - y^{(i)}) \frac{1}{1 - h(x^{(i)})} \right) \frac{\partial}{\partial \theta_j} (h(x^{(i)})) \\ = \left(\frac{y^{(i)}}{g(\theta^T x^{(i)})} - (1 - y^{(i)}) \frac{1}{1 - g(\theta^T x^{(i)})} \right) \frac{\partial}{\partial \theta_j} (g(\theta^T x^{(i)})) \\ = \left(\frac{y^{(i)}}{g(\theta^T x^{(i)})} - (1 - y^{(i)}) \frac{1}{1 - g(\theta^T x^{(i)})} \right) g(\theta^T x^{(i)}) (1 - g(\theta^T x^{(i)})) \frac{\partial \theta^T x^{(i)}}{\partial \theta_j} \\ = (y^{(i)} (1 - g(\theta^T x^{(i)})) - (1 - y^{(i)}) g(\theta^T x^{(i)})) x_j \\ = (y^{(i)} - h_{\theta}(x^{(i)})) x_j$$

从而迭代θ至收敛即可：

$$\theta_j := \theta_j + \alpha (y^{(i)} - h_{\theta}(x^{(i)})) x_j^{(i)}$$

3 模型评估

对于LR分类模型的评估，常用AUC来评估，关于AUC的更多定义与介绍，可见参考文献2，在此只介绍一种极简单的计算与理解方法。

对于下图的分类：

数据 ⁽ⁱ⁾	分类 ⁽ⁱ⁾
10 ⁽ⁱ⁾	1 ⁽ⁱ⁾
9 ⁽ⁱ⁾	1 ⁽ⁱ⁾
8 ⁽ⁱ⁾	1 ⁽ⁱ⁾
7 ⁽ⁱ⁾	1 ⁽ⁱ⁾
6 ⁽ⁱ⁾	1 ⁽ⁱ⁾
5 ⁽ⁱ⁾	0 ⁽ⁱ⁾
4 ⁽ⁱ⁾	0 ⁽ⁱ⁾
3 ⁽ⁱ⁾	0 ⁽ⁱ⁾
2 ⁽ⁱ⁾	0 ⁽ⁱ⁾
1 ⁽ⁱ⁾	0 ⁽ⁱ⁾

数据 ⁽ⁱ⁾	预测 ⁽ⁱ⁾	分类 ⁽ⁱ⁾
10 ⁽ⁱ⁾	1 ⁽ⁱ⁾	1 ⁽ⁱ⁾
9 ⁽ⁱ⁾	0.9 ⁽ⁱ⁾	1 ⁽ⁱ⁾
8 ⁽ⁱ⁾	0.8 ⁽ⁱ⁾	1 ⁽ⁱ⁾
7 ⁽ⁱ⁾	0.7 ⁽ⁱ⁾	1 ⁽ⁱ⁾
6 ⁽ⁱ⁾	0.5 ⁽ⁱ⁾	0 ⁽ⁱ⁾
5 ⁽ⁱ⁾	0.6 ⁽ⁱ⁾	1 ⁽ⁱ⁾
4 ⁽ⁱ⁾	0.4 ⁽ⁱ⁾	0 ⁽ⁱ⁾
3 ⁽ⁱ⁾	0.3 ⁽ⁱ⁾	0 ⁽ⁱ⁾
2 ⁽ⁱ⁾	0.2 ⁽ⁱ⁾	0 ⁽ⁱ⁾
1 ⁽ⁱ⁾	0.1 ⁽ⁱ⁾	0 ⁽ⁱ⁾

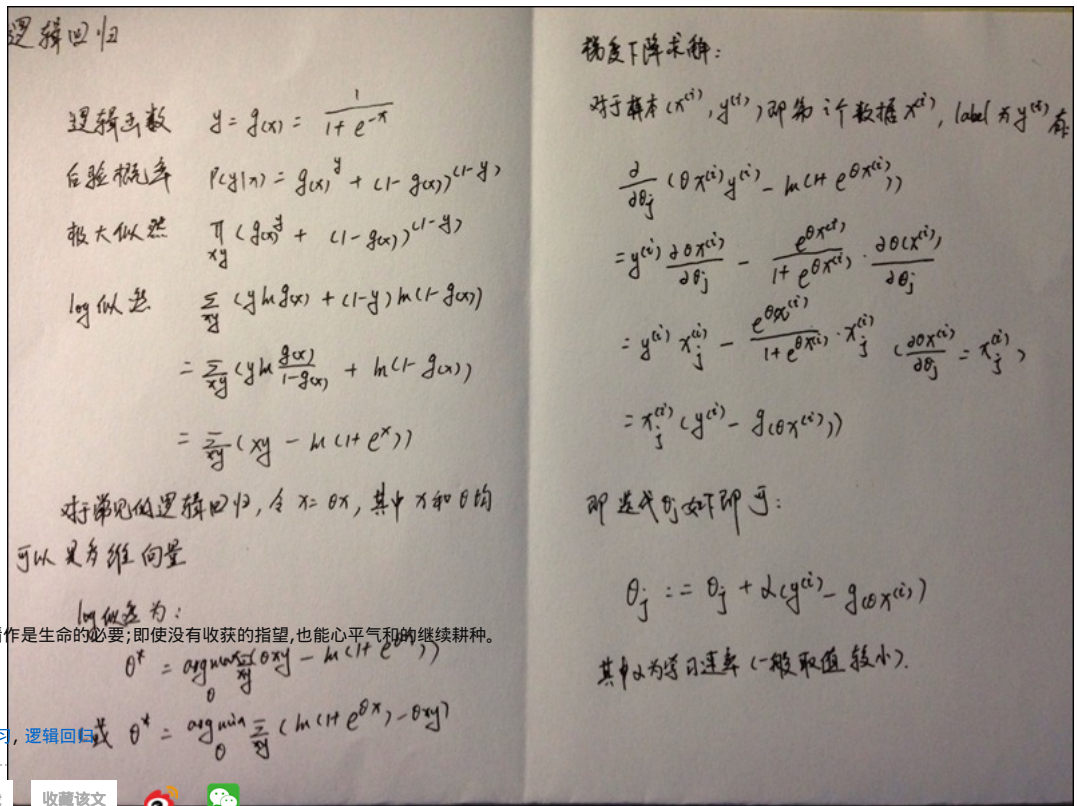
数据 ⁽ⁱ⁾	预测 ⁽ⁱ⁾	分类 ⁽ⁱ⁾
10 ⁽ⁱ⁾	0.1 ⁽ⁱ⁾	0 ⁽ⁱ⁾
9 ⁽ⁱ⁾	0.9 ⁽ⁱ⁾	1 ⁽ⁱ⁾
8 ⁽ⁱ⁾	0.8 ⁽ⁱ⁾	1 ⁽ⁱ⁾
7 ⁽ⁱ⁾	0.7 ⁽ⁱ⁾	1 ⁽ⁱ⁾
6 ⁽ⁱ⁾	0.6 ⁽ⁱ⁾	1 ⁽ⁱ⁾
5 ⁽ⁱ⁾	0.5 ⁽ⁱ⁾	0 ⁽ⁱ⁾
4 ⁽ⁱ⁾	0.4 ⁽ⁱ⁾	0 ⁽ⁱ⁾
3 ⁽ⁱ⁾	0.3 ⁽ⁱ⁾	0 ⁽ⁱ⁾
2 ⁽ⁱ⁾	0.2 ⁽ⁱ⁾	0 ⁽ⁱ⁾
1 ⁽ⁱ⁾	1 ⁽ⁱ⁾	1 ⁽ⁱ⁾

对于训练集的分类，训练方法1和训练方法2分类正确率都为80%，但明显可以感觉到训练方法1要比训练方法2好。因为训练方法1中，5和6两数据分类错误，但这两个数据很近，而训练方法2中，将10和1两个数据分类错误，但这两个数据均离分类面较远。

AUC正是衡量分类正确度的方法，将训练集中的label看两类{0, 1}的分类问题，分类目标是将预测结果尽量将两者分开。将每个0和1看成一个pair关系，团中的训练集共有!pair关系，只有将所有pair关系一至时，分类结果才是最好的，而auc为1。在训练方法1中，与10相关的pair关系完全正确，同样9、8、7的pair关系也完全正确，但对于6，其(5)关系错误，而与4、3、2、1的关系正确，故其auc为(25-1)/25=0.96；对于分类方法2，其6、7、8、9的pair关系，均有一个错误，即(6,1)、(7,1)、(8,1)、(9,1)，对于其正任何数据点的pair关系，都错误，即(10,1)、(10,2)、(10,3)、(10,4)、(10,5)，故方法2的auc为(25-4-5)/25=0.64，因而正如直观所见，分类方法1要优于分类方法2。

4 演算手稿

附演算手稿如下：



只能永远把艰辛的劳动看作是生命的必要;即使没有收获的指望,也能心平气和的继续耕种。

分类: 数据挖掘与推荐

标签: 数据挖掘, 机器学习, 逻辑回归

好文要顶

关注我

收藏该文



文墨

关注 -

12

+ 加关注

粉丝

50

1. <http://www.cnblogs.com/quelei/archive/2013/05/23/3095747.html>

2. 上一篇: 简单的基于地理图片的旅行路线还原

3. 下一篇: 谱聚类算法(Spectral Clustering)优化与扩展

1

posted @ 2013-11-25 11:56 文墨 阅读(16768) 评论(

评论列表

#1楼 2013-11-26 14:04 acfun

大香肠给跪了,数学全还给老师了

支持(

#2楼 2013-11-27 10:01 寻风问雨

没太懂楼主所谓的机器学习定义是什么。

支持(

#3楼 2014-04-08 11:19 CrazySpider

内容很好,不过有几个地方貌似有误:

(1) 图2逻辑曲线图的横坐标,应该是在 $z=0$ 的点取得 $g(z)=0.5$

(2) 方程(4)中第一个等号右边的式子分母部分, e 的指数符号貌似应该是正号

(3) 采用最大似然估计的时候,用的是最大似然函数,那么用梯度法求解目标优化问题的时候,就应该是求解目标函数的最大值而不是最小值,因此所采用的算法应该是梯度上升法(或者对目标函数取负,然后用梯度下降法)

支持(

#4楼 2014-08-12 18:34 CodeBreaker

@ CrazySpider

补充第(3)点就是,原文在求梯度的时候把求和符号去掉了。。。

支持(

#5楼 2016-06-11 16:11 似水流年啊啊啊

您好,您有逻辑回归的代码吗

支持(

#6楼 2016-08-02 17:29 萌悦

后验概率应该用乘积的形式,加法的形式导不出后面的结果

支持(

#7楼[楼主] 2016-10-12 11:01 文墨

@ 似水流年啊啊啊

这个网上到处都有的

支持(

#8楼 2017-03-16 11:14 air_fans

看到一堆公式就头大了,能否用通俗的语言描述原理?

支持(

刷新评论 刷新页

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，访问[网站首页](#)。

【推荐】50万行VC++源码：大型组态工控、电力仿真CAD与GIS源码库

【推荐】中铁、中石油等大型企业的复杂报表解决方案

【活动】阿里云海外云服务全面降价助力企业全球布局

【实用】40+篇云服务器操作及运维基础知识！



最新IT新闻：

- 重构之十六字心法
 - 巴菲特清仓IBM？一位IBM人员反思：战略领先十年，模式落后百年
 - Visual Studio Code的PowerShell扩展升至1.0版本
 - 摩拜单车黑科技加身 再不怕车筐被贴小广告
 - 微软发布MySQL/PostgreSQL Azure服务与Cosmos数据库
- » [更多新闻...](#)



最新知识库文章：

- 唱吧DevOps的落地，微服务CI/CD的范本技术解读
 - 程序员，如何从平庸走向理想？
 - 我为什么鼓励工程师写blog
 - 怎么轻松学习JavaScript
 - 如何打好前端游击战
- » [更多知识库文章...](#)