

机器学习统计

对数损失函数是如何度量损失的？

对于损失函数，像用平方损失函数  $L(Y, f(X)) = (Y - f(X))^2$  或者绝对值损失函数都十分直观，但是对数损失函数  $L(Y, P(Y|X))$ ...显示全部

关注问题

写回答

2 条评论

分享

邀请回答

6 个回答

默认排序



王赞 Maigo

日语、语言、机器学习 等 6 个话题的优秀回答者

66 人赞同了该回答

对数损失是用于最大似然估计的。

一组参数在一堆数据下的似然值，等于每一条数据在这组参数下的条件概率之积。

而损失函数一般是每条数据的损失之和，为了把积变为和，就取了对数。

再加个负号是为了让最大似然值和最小损失对应起来。

编辑于 2018-03-10

66

8 条评论

分享

收藏

感谢



知乎用户

csuldw.com

72 人赞同了该回答

说到log损失，情不自禁地让我想到了逻辑回归的推导。

有些人可能觉得逻辑回归的损失函数就是平方损失，其实并不是。平方损失函数是线性回归在假设样本是高斯分布的条件下推导得到的（为什么假设成高斯分布呢？其实这里隐藏了一个小知识点，就是中心极限定理，可以参考【Central limit theorem】，而逻辑回归得到的并不是平方损失。在逻辑回归的推导中，它假设样本服从伯努利分布（0-1分布），然后求得满足该分布的似然函数，接着取对数等等。逻辑回归并没有极大化似然函数求极值，而是把极大化当做是一种思想，进而推导出它的经验风险函数为：最小化负的似然函数（即  $\max F(y, f(x)) \rightarrow \min -F(y, f(x))$ ）。从损失函数的视角来看，它就成了log损失函数了。

log损失函数的标准形式：

$$L(Y, P(Y|X)) = -\log P(Y|X)$$

刚刚说到，对数是用来做极大似然估计的，图的是计算方便，因为在MLE中，直接求导比较困难，所以通常都是先取对数再寻找极值点。损失函数  $L(Y, P(Y|X))$  表达的是样本X在分类Y的情况下，使概率  $P(Y|X)$  达到最大，即  $\max P(Y|X)$ 。而log是单调递增函数，所以  $\log P(Y|X)$  也会达到最大值，因此前面加符号之后，最大化  $P(Y|X)$  就等同于最小化L。

逻辑回归的  $P(Y=y|x)$  表达式如下（为了将类别标签y统一为1和0，下面将表达式分开表示）：

$$P(Y = y|x) = \begin{cases} h_\theta(x) = g(f(x)) = \frac{1}{1 + \exp\{-f(x)\}} & , y = 1 \\ 1 - h_\theta(x) = 1 - g(f(x)) = \frac{1}{1 + \exp\{f(x)\}} & , y = 0 \end{cases}$$

将它带入到上式，通过微分可得，在给定数据下，找到最有可能导致这种分布的参数值；或者说什么样的参数才能使我们观测到目前这组数据的概率最大。而log是单调递增函数，所以  $\log P(Y|X)$  也会达到最大值，因此前面加符号之后，最大化  $P(Y|X)$  就等同于最小化L。

72

6 条评论

收藏

感谢

收起

相关推荐

刘看山 · 知乎指南 · 知乎协议 · 知乎开放平台 · 知乎隐私政策 · 知乎广告 · 知乎招聘 · 侵权举报 · 网上有害信息举报专区 · 违法和不良信息举报：010-82717171 · 儿童色情信息举报专区 · 联系我们 · © 2018 知乎

机器学习中的目标函数、损失函数有什么区别？ 21 个回答

关于二元Logistics回归的损失函数？ 1 个回答

在统计学中为什么要对变量取对数？ 1 个回答

逻辑回归损失函数为什么使用交叉熵而不是最小二乘法？ 6 个回答

吴军：带你考...  
吴军  
共 24 节课

和吴军、刘长...  
创业  
★★★★★ 1

机器学习：实...  
Drew Conway

https://www.zhihu.com/question/27126057

1/3

## 相关推荐

刘看山 · 知乎指南 · 知乎协议 · 知乎隐私政策

申请开通知乎机构号

侵权举报 · 网上有害信息举报专区

违法和不良信息举报: 010-8271

儿童色情信息举报专区

联系我们 © 2018 知乎

$$L(y, P(Y = y|x)) = \begin{cases} \log(1 + \exp\{-f(x)\}) & , y = 1 \\ \log(1 + \exp\{f(x)\}) & , y = 0 \end{cases}$$

逻辑回归最后得到的目标式子如下：

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right]$$

值得说明的是：之所以有人认为逻辑回归是平方损失，是因为在使用梯度下降来求最优解的时候，它的迭代式子与平方损失求导后的式子非常相似，从而给人一种错觉了。

以上内容纯个人理解，如有纰漏，还望大家指正！

编辑于 2016-09-19



知乎用户

某大学大学生

12 人赞同了该回答

$L(Y, P(Y|X))$ 这个对数损失函数的意思是指分类为Y的情况下，使 $P(Y|X)$ 达到最大。你要想一下，有一些模型是用最大概率的分类来做预测的，而Y是代表分类为正确的分类，而 $P(Y|X)$ 则是代表正确分类的概率，那对数取反是不是 $P(Y|X)$ 越大，损失函数就越小？

发布于 2015-01-22

▲ 12 ▼ ● 7 条评论 ➦ 分享 ★ 收藏 ♥ 感谢



DomainAdaptation

以统一的视角重构机器学习

6 人赞同了该回答

问题中提到的对数损失函数的形式，在李航的《统计学习方法》第一章中见到过。

我们可以从损失的本质去回答这个问题。损失是真实模型与假设模型之间差异的度量。这里的模型有可能是直接用于回归或分类的函数模型，也就是说学出来的函数可以直接用来做回归或分类任务(分类任务需要另外套上一个符号函数)，也有可能是一个条件概率分布模型(做回归与分类任务均可，关键在于假定分布的形式)。

平方损失函数之所以直观，是因为它直接度量了真实回归模型与假设回归模型之间的差异；而对数损失则是度量了真实条件概率分布与假定条件概率分布之间的差异，而这里的差异度量用的是KL散度。

这个问题其实延伸出来，会牵扯到概率模型与非概率模型的问题。非概率模型的学习通过预先选定损失函数如平方损失，hinge损失，指数损失等0-1损失的替代品，然后通过最小化平均损失的形式来学习函数模型；而概率模型的学习则是预先选定条件分布的形式，然后通过最小化某种概率分布距离的形式来学习分布模型。

发布于 2017-12-29

▲ 6 ▼ ● 添加评论 ➦ 分享 ★ 收藏 ♥ 感谢



Dylan

IT工程师 机器学习学者

4 人赞同了该回答

这里用了似然的思想在里面！ $P(Y|X)$ 表示当前模型下样本X得到Y的概率，一定意义上也是样本和目标值的接近程度。用到损失函数上，取个反！这个概率越大，样本和目标值越接近，损失越小，反之亦然！

发布于 2016-02-01

▲ 4 ▼ ▲ 72 ● 6 条评论 ★ 收藏 ♥ 感谢

收起

## 相关问题



下载知乎客户端  
与世界分享知识、经验



**codlife**

Spark Contributor，数据工作者

1 人赞同了该回答

针对rank 1 的解答，有一点不太赞同的地方，他说LR的推导过程中假设样本服从伯努利分布，但是：不是说当二项分布的N充分大，并且p 不接近1或0，那么该二项分布就接近高斯分布，这样的话，基本也符合高斯分布了。 如果使用MSE作Loss会导致损失函数非凸。还请牛人指正。参考：[逻辑回归 - 博客频道 - CSDN.NET](#)

编辑于 2017-02-28

▲ 1 ▼

添加评论

分享

★ 收藏

♥ 感谢

✎ 写回答

1 个回答被折叠（为什么？）

相关推荐

刘看山 · 知乎指南 · 知乎协议 · 广告  
申请开通知乎机构号  
侵权举报 · 网上有害信息举报专区  
违法和不良信息举报：010-8271  
儿童色情信息举报专区  
联系我们 © 2018 知乎