

CSDN

博客学院下载图文课TinyMind论坛APP问答商城VIP会员活动招聘ITeyeGitChat

搜博主文章

写博客发Chat

原

一文搞懂交叉熵在机器学习中的使用，透彻理解交叉熵背后的直觉

2018年01月25日 16:45:50 史丹利复合田 阅读数：18406 标签：深度学习交叉熵优化函数更多

版权声明：本文为博主原创文章，转载请注明出处。https://blog.csdn.net/tsycnh/article/details/79163834

关于交叉熵在loss函数中使用的理解

交叉熵（cross entropy）是深度学习中常用的一个概念，一般用来求目标与预测值之间的差距。以前做一些分类问题的时，用起来也比较方便。最近开始研究起对抗生成网络（GANs），用到了交叉熵，发现自己对交叉熵的理解有些模糊，不如下相关知识点，才算透彻的理解了，特地记录下来，以便日后查阅。

信息论

交叉熵是信息论中的一个概念，要想了解交叉熵的本质，需要先从最基本的概念讲起。

1 信息量

首先是信息量。假设我们听到了两件事，分别如下：
事件A：巴西队进入了2018世界杯决赛圈。
事件B：中国队进入了2018世界杯决赛圈。
仅凭直觉来说，显而易见事件B的信息量比事件A的信息量要大。究其原因，是因为事件A发生的概率很大，事件B发生的概率很小。所以当越不可能的我们获取到的信息量就越大。越可能发生的事件发生了，我们获取到的信息量就越小。那么信息量应该和事件发生的概率有关。

假设X是一个离散型随机变量，其取值集合为χ,概率分布函数 $p(x) = Pr(X = x), x \in \chi$,则定义事件 $X = x_0$ 的信息量为：
$$I(x_0) = -\log(p(x_0))$$

由于是概率所以 $p(x_0)$ 的取值范围是[0, 1],绘制为图形如下：

可见该函数符合我们对信息量的直觉

2 熵

考虑另一个问题，对于某个事件，有n种可能性，每一种可能性都有一个概率 $p(x_i)$
这样就可以计算出某一种可能性的信息量。举一个例子，假设你拿出了你的电脑，按下开关，会有三种可能性，下表列出了每一种可能的概率及其对应

序号	事件	概率p	信息量I
A	电脑正常开机	0.7	$-\log(p(A))=0.36$
B	电脑无法开机	0.2	$-\log(p(B))=1.61$
C	电脑爆炸了	0.1	$-\log(p(C))=2.30$

注：文中的对数均为自然对数

我们现在有了信息量的定义，而熵用来表示所有信息量的期望，即：

黑马程序员 美甲培训学校 自考网官网 赶集网租房

登录 注册

https://blog.csdn.net/tsycnh/article/details/79163834 1/9

$$H(X) = - \sum_{i=1}^n p(x_i) \log(p(x_i))$$

其中n代表所有的n种可能性，所以上面的问题结果就是

$$\begin{aligned} H(X) &= -[p(A)\log(p(A)) + p(B)\log(p(B)) + p(C)\log(p(C))] \\ &= 0.7 \times 0.36 + 0.2 \times 1.61 + 0.1 \times 2.30 \\ &= 0.804 \end{aligned}$$

然而有一类比较特殊的问题，比如投掷硬币只有两种可能，字朝上或花朝上。买彩票只有两种可能，中奖或不中奖。我们称之为0-1分布问题（二项分布对于这类问题，熵的计算方法可以简化为如下算式：

$$\begin{aligned} H(X) &= - \sum_{i=1}^n p(x_i) \log(p(x_i)) \\ &= -p(x) \log(p(x)) - (1-p(x)) \log(1-p(x)) \end{aligned}$$

3 相对熵（KL散度）

相对熵又称KL散度,如果我们对于同一个随机变量 x 有两个单独的概率分布 $P(x)$ 和 $Q(x)$ ，我们可以使用 KL 散度（Kullback-Leibler divergence）（KL）divergence）分布的差异

维基百科对相对熵的定义

In the context of machine learning, DKL(P||Q) is often called the information gain achieved if P is used instead of Q.

即如果用P来描述目标问题，而不是用Q来描述目标问题，得到的信息增量。

在机器学习中，P往往用来表示样本的真实分布，比如[1,0,0]表示当前样本属于第一类。Q用来表示模型所预测的分布，比如[0.7,0.2,0.1]直观的理解就是如果用P来描述样本，那么就非常完美。而用Q来描述样本，虽然可以大致描述，但是不是那么的完美，信息量不足，需要额外的一些能达到和P一样完美的描述。如果我们的Q通过反复训练，也能完美的描述样本，那么就不再需要额外的“信息增量”，Q等价于P。

KL散度的计算公式：

$$D_{KL}(p||q) = \sum_{i=1}^n p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right)$$

n为事件的所有可能性。

D_{KL} 的值越小，表示q分布和p分布越接近

4 交叉熵

对式3.1变形可以得到：

$$\begin{aligned} D_{KL}(p||q) &= \sum_{i=1}^n p(x_i) \log(p(x_i)) - \sum_{i=1}^n p(x_i) \log(q(x_i)) \\ &= -H(p(x)) + \left[- \sum_{i=1}^n p(x_i) \log(q(x_i))\right] \end{aligned}$$

等式的前一部分恰巧就是p的熵，等式的后一部分，就是交叉熵：

$$H(p, q) = - \sum_{i=1}^n p(x_i) \log(q(x_i))$$

在机器学习中，我们需要评估label和predicts之间的差距，使用KL散度刚刚好，即 $D_{KL}(y||\hat{y})$ ，由于KL散度中的前一部分 $-H(y)$ 不变，故在优化过程注交叉熵就可以了。所以一般在机器学习中直接用交叉熵做loss，评估模型。

机器学习中交叉熵的应用

1 为什么要用交叉熵做loss函数？

在线性回归问题中，常常使用MSE（Mean Squared Error）作为loss函数，比如：

$$loss = \frac{1}{2m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

这里的m表示m个样本的，loss为m个样本的loss均值。

MSE在线性回归问题中比较好用，那么在逻辑分类问题中还是如此么？

2 交叉熵在单分类问题中的使用

这里的单类别是指，每一张图像样本只能有一个类别，比如只能是狗或只能是猫。
交叉熵在单分类问题上基本是标配的方法

$$loss = - \sum_{i=1}^n y_i \log(\hat{y}_i)$$

上式为一张样本的loss计算方法。式2.1中n代表着n种类别。
举例说明,比如有如下样本



对应的标签和预测值

*	猫	青蛙	老鼠
Label	0	1	0
Pred	0.3	0.6	0.1

那么

$$\begin{aligned} loss &= -(0 \times \log(0.3) + 1 \times \log(0.6) + 0 \times \log(0.1)) \\ &= -\log(0.6) \end{aligned}$$

对应一个batch的loss就是

$$loss = - \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n y_{ji} \log(\hat{y}_{ji})$$

m为当前batch的样本数

3 交叉熵在多分类问题中的使用

这里的多类别是指，每一张图像样本可以有多个类别，比如同时包含一只猫和一只狗
和单分类问题的标签不同，多分类的标签是n-hot。
比如下面这张样本图，即有青蛙，又有老鼠，所以是一个多分类问题



对应的标签和预测值

*	猫	青蛙	老鼠
Label	0	1	1
Pred	0.1	0.7	0.8

值得注意的是，这里的Pred不再是通过softmax计算的，这里采用的是sigmoid。将每一个节点的输出归一化到[0,1]之间。所有Pred值的和也不再为1就是每一个Label都是独立分布的，相互之间没有影响。所以交叉熵在这里是单独对每一个节点进行计算，每一个节点只有两种可能值，所以是一个二说过对于二项分布这种特殊的分布，熵的计算可以进行简化。

同样的，交叉熵的计算也可以简化，即

$$loss = -ylog(\hat{y}) - (1 - y)log(1 - \hat{y})$$

注意，上式只是针对一个节点的计算公式。这一点一定要和单分类loss区分开来。

例子中可以计算为：

$$\begin{aligned} loss_{猫} &= -0 \times \log(0.1) - (1 - 0)\log(1 - 0.1) = -\log(0.9) \\ loss_{蛙} &= -1 \times \log(0.7) - (1 - 1)\log(1 - 0.7) = -\log(0.7) \\ loss_{鼠} &= -1 \times \log(0.8) - (1 - 1)\log(1 - 0.8) = -\log(0.8) \end{aligned}$$

单张样本的loss即为 $loss = loss_{猫} + loss_{蛙} + loss_{鼠}$
每一个batch的loss就是：

$$loss = \sum_{j=1}^m \sum_{i=1}^n -y_{ji}log(\hat{y}_{ji}) - (1 - y_{ji})log(1 - \hat{y}_{ji})$$

式中m为当前batch中的样本量，n为类别数。

总结

路漫漫，要学的东西还有很多啊。

参考：

<https://www.zhihu.com/question/65288314/answer/244557337>
https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence
<https://jamesmccaffrey.wordpress.com/2013/11/05/why-you-should-use-cross-entropy-error-instead-of-classification-error-or-mean-squared-error-for-neural-network-clng/>

- 小小小小小小小小菜鸟：这个不错 (09-28 14:29 #16楼)
- qq_27602353：通俗易懂！ (09-20 22:46 #15楼)
- 逍遥如我：真滴很优秀 (09-20 17:41 #14楼)

查看 49 条热评

深度学习用到的代价函数-----交叉熵

交叉熵代价函数（ Cross-entropy cost function ）是用来衡量神经网络（ ANN ）的预测值与实际值的一种方式。与二次代价函数相比，它能更有效地...

1677

交叉熵为何能作损失函数

在很多二分类问题中，特别是正负样本不均衡的分类问题中，常使用交叉熵作为loss对模型的参数求梯度进行更新，那为何交叉熵能作为损失函数呢，我...

1.8万

 有哪些可以免费试用一年左右的云服务器

百度广告

交叉验证

交叉验证

194

机器学习中的熵和 loss function

什么是熵？熵是由热力学领域首次提出的。在机器学习中，一个事件出现的概率越小，那么当这个事件发生时，它所代表的信息量越大。 对于一个随...

233

机器学习 - 交叉熵Cross Entropy

机器学习 - 交叉熵Cross Entropy 1. 二值交叉熵 Binary Cross Entropy 假设训练数据 $D=\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$ $D=\{...$

2132

交叉熵

文章转自：http://blog.csdn.net/rtygbwwwerr/article/details/50778098 交叉熵（ Cross-Entropy ）交叉熵是一个在ML领域经常会被提...

1.5万

这传奇你卸载算我输！爆率9.8，有充值入口我跪键盘！

含玩游戏·顶新

交叉熵的解决代价函数为二次函数导致学习慢问题(S型激..._CSDN博客

TensorFlow 笔记-06-神经网络优化-交叉熵,学习率 交叉熵交叉熵(Cross Entropy):表示两个概率分布之间的距离,交叉熵越大,两个概率分布距离越远,两个概...

(Tensorflow之四)激活函数、交叉熵及优化器 - CSDN博客

在深度学习中,可以将交叉熵作为Loss函数,目标减少p与q之间的距离; 三、优化器 3.1 梯度下降法GradientDescentOptimizer 原理:若函数f(x,y,z)可导,则(af/...

关于交叉熵

在信息论中可以根据香农熵估计一个字符串中每个符号 S_i 编码所需的最小位数，前提是每个字符出现的概率 P_i 是已知的。 $H=-\sum (P_i \times \log_2(P_i))$ $H=-...$

1160

交叉熵（ Cross-Entropy ）

交叉熵（ Cross-Entropy ）交叉熵是一个在ML领域经常会被提到的名词。在这篇文章里将对这个概念进行详细的分析。 1.什么是信息量？ 假设XX是一个...

8.3万

简单易懂的softmax交叉熵损失函数求导

来写一个softmax求导的推导过程，不仅可以给自己理清思路，还可以造福大众，岂不美哉~ softmax经常被添加在分类任务的神经网络中的输出层，神经...

1.5万

交叉熵代价函数（作用及公式推导）

交叉熵代价函数（ Cross-entropy cost function ）是用来衡量神经网络（ ANN ）的输出值与实际值的一种方式。与二次代价函数（ Quadratic cost func...

6.4万

文章热词

深度学习图片分类方法 深度学习 目标标注 深度学习算法类别 多智能体深度学习 深度学习分布式模型

相关热词

一文hmm 一文深度学习 一文看懂 一文回调函数 一文编码

博主推荐



luchi007

关注

74篇文章



pku_zzy

关注

166篇文章



亚当-adam

关注

55篇文章

交叉熵(Cross Entropy) - CSDN博客

交叉熵应用到机器学习问题的简要介绍... 交叉熵应用到机器学习问题的简要介绍 交叉熵可以用于定义机器学习问题 and 最优化问题的损失函数。我们可以假...

交叉熵方法求解组合优化问题

交叉熵方法是一个自适应的收敛优化方法,对解决优化问题效果非常的好... 交叉熵方法是一个自适应的收敛优化方法,对解决优化问题效果非常的好 交叉熵...

深度学习中的交叉熵误差原理

深度学习中的交叉熵误差原理,深度学习中的交叉熵误差原理,深度学习中的交叉熵误差原理

神经网络学习(八)优化方法:交叉熵代价函数 - CSDN博客

下面继续进行学习,学习其他的优化方法。 问题 我们通常是在犯比较严重的错误时...如果我们使用均方误差代价函数或者上面的交叉熵函数(2) 要...

机器学习 | 交叉熵 - CSDN博客

为分类任务中,常用交叉熵来刻画损失函数,为什么可以这样使用呢?下面将一步步的揭开面纱。 1、数学背景定义:在信息论中,交叉熵是表示两个概率分布p,...

交叉熵为何能作损失函数 - CSDN博客

在很多二分类问题中,特别是正负样本不平衡的分类问题中,常使用交叉熵作为loss对模型的参数求梯度进行更新,那为何交叉熵能作为损失函数呢,我也是带...

简单的交叉熵损失函数，你真的懂了吗？

个人网站：红色石头的机器学习之路 CSDN博客：红色石头的专栏 知乎：红色石头 微博：RedstoneWill的微博 GitHub：RedstoneWill的GitHub...

数据挖掘笔记-特征选择-期望交叉熵

期望交叉熵也称为KL距离，反映的是文本类别的概率分布和在出现了某个特征的条件下文本类别的概率分布之间的距离，具体公式表示如下 其中， P(t)表...

为什么要用交叉熵作为代价函数

对于大多数人来说，犯错是一件让人很不开心的事情。但反过来想，犯错可以让我们意识到自己的不足，然后我们很快就学会下次不能再犯错了。犯的错...

云南女孩揭秘茶叶行业潜规则，常喝茶的人一定得知道！

讯酷·爆料

【联系】二项分布的对数似然函数与交叉熵 (cross entropy) 损失函数

1. 二项分布二项分布也叫 0-1 分布，如随机变量 x 服从二项分布，关于参数 μ ($0 \leq \mu \leq 1$)，其取值 1 和取 0 的概率如下： $\{p(x=1|\mu)=...$

交叉熵及其使用

原文链接：<https://blog.csdn.net/tsycnh/article/details/79163834> 最近在看交叉熵，读到这一篇很有帮助，转载学习，谢谢原文博主的讲解！ 关于交叉...

2线性分类器基本原理-2.3线性分类器之SoftMax和交叉熵损失(Cross-Entropy)

图像分类器模型的整体结构：交叉熵 (Cross-Entropy) 损失 和 SoftMax SVM是最常用的两个分类器之一，而另一个就是Softmax分类器，它的损失函数...

交叉熵以及为什么用交叉熵作为代价函数

作者：知乎用户链接：<https://www.zhihu.com/question/41252833/answer/108777563>来源：知乎著作权归作者所有。商业转载请联系作者获得授权，...

神经网络与深度学习笔记（四）为什么用交叉熵代替二次代价函数

1、为什么不用 二次代价函数 我们用的 loss 方程是 a 是神经元的输出，其中 $a = \sigma(z)$ ， $z = wx + b$ 使用链式法则求权重和偏置的偏导数有：可知，偏导...



机械键盘十大知名品牌是什么

百度广告

理解交叉熵

交叉熵公式 $H(p,q)=$ 交叉熵是用于度量两个分部距离的 如度量两片文章相似度时， i 代表第*i*个单词， $p(i)$ 代表这个单词出现频率 如将其作为神经网络误差函...

<https://blog.csdn.net/tsycnh/article/details/79163834>

6/9

为什么在训练神经网络时候比较适合使用交叉熵错误率，而不是分类错误率或是均方差

原文：Why You Should Use Cross-Entropy Error Instead Of Classification Error Or Mean Squared Error For ...

神经网络反向传播算法的推导

神经网络反向传播算法的推导

机器学习经典损失函数之交叉熵和均方差

损失函数是用来评估预测模型效果的，即model的预测值与实真实值的差距。不用的机器学习模型有不同的损失函数来评估。本篇博文主要讲用于分类问...

机器学习使用交叉熵作为损失函数的原因

在机器学习中的分类算法中，我们总是最小化交叉熵，因为交叉熵越低，就证明由算法所产生的策略最接近最优策略，也间接证明我们算法所算出的非...

《开国大典》一公斤纯银纪念币，限时抢藏价980元。

如意·顶新

逻辑回归算法之交叉熵函数理解

逻辑回归算法之交叉熵函数理解

理解交叉熵(cross_entropy)作为损失函数在神经网络中的作用

交叉熵的作用通过神经网络解决多分类问题时，最常用的一种方式就是在最后一层设置n个输出节点，无论在浅层神经网络还是在CNN中都是如此，比如...

交叉熵、相对熵及KL散度通俗理解

原文转载自<http://blog.csdn.net/u012162613/article/details/44239919>本文是《Neural networks and deep learning》概...

交叉熵代价函数

本文是《Neural networks and deep learning》概览 中第三章的一部分，讲machine learning算法中用得很多的交叉熵代价函数。1.从方差代价函数说起代...

交叉熵是什么，能干什么用，

在做tensorflow对手写mnist字体识别中，用到了交叉熵来定义损失函数的方法，看着别人用这个，但是又不知道原理，心里特别的不踏实，就特地查了一...



内地人来香港银行开户攻略

百度广告

多分类问题中的交叉熵

熵的本质是香农信息量()的期望。现有关于样本集的2个概率分布p和q，其中p为真实分布，q非真实分布。按照真实分布p来衡量识别一个样本的所需要的...

交叉熵 机器学习模型loss指标

华夏35度 Data Mining http://www.cnblogs.com/zhangchaoyang/articles/2655785.html?utm_source=tuicool...

GAN学习笔记（1）：GAN综述

GAN学习笔记（1）：GAN综述，针对GAN的基本模型，数学原理，衍生模型，应用领域等，根据大量论文、知乎、CSDN博客网络资料整理...

特征选择方法之期望交叉熵

本文转自：http://blog.csdn.net/fighting_one_piece/article/details/38562183 期望交叉熵也称为KL距离，反映的是文本类别的概率分...

机器学习（四）从信息论交叉熵的角度看softmax/逻辑回归损失

机器学习中会常见到softmaxLoss，逻辑回归损失（或者叫交叉熵损失），这两种损失的来源可以由两方面考虑，一方面可以看做是来源于概率论中的极...

陈小春：我在这游戏挂机一晚，打赢了古天乐，挂机赢笑了！

贪玩游戏·顶新

深度学习---之softmax与交叉熵

交叉熵到底是如何来定义损失函数的呢？首先看下面的公式：不用完全对着上图公式看，因为我的解释没有加上log和-号，你只要看懂文字，加log和-号...

为什么在神经网络中要使用交叉熵而不是均方差作为损失函数

  623

https://jamesmccaffrey.wordpress.com/2013/11/05/why-you-should-use-cross-entropy-error-instead-of-cl...

关于交叉熵的个人理解

  40

在信息论和机器学习中，常常会聊到交叉熵这个概念，用这个专有名词来计算预测结果与实际结果的差距，简单地说就是判断学习的好与坏。那为了方...



下载

深度学习中的交叉熵误差原理

12-07

深度学习中的交叉熵误差原理，深度学习中的交叉熵误差原理，深度学习中的交叉熵误差原理

信息熵 交叉熵 相对熵 条件熵

  860



根据香农公式，信息量等于 $\log(1/p)\log(\frac{1}{p})$ ；交叉熵常作为机器学习中的损失函数。信息熵 熵的本质是信息量的期望： $H(p)=\sum p(i)*\log(1/p(...$



为什么林肯suv这么便宜



百度广告

交叉熵损失函数

  2.7万

一、香农熵 香农熵 1948 年，香农提出了“信息熵”(shāng) 的概念，才解决了对信息的量化度量问题。一条信息的信息量大小和它的不确定性有直接的关...

有关熵的几个概念 及 最大似然和交叉熵的一致性

  3599

随机事件的信息量 为了描述一个随机事件的信息量，定义了自信息。自信息表示不确定性减少的程度。 一个事件确定发生，是没有信息量的；...



史丹利复合田

关注

原创

粉丝

喜欢

评论

52

165

120

129

等级： 博客 5

访问：19万+

积分：2489

排名：2万+

勋章： 



celine价格



最新文章

直观的理解tensorflow中的tf.tile()函数

微信聊天记录导出工具WeChatExporter开源啦！

关于tensorflow中的softmax_cross_entropy_with_logits_v2函数的区别

Python 小技巧

Mac下如何建立双击可执行的脚本

博主专栏



Keras入门课

阅读量：12940 7 篇



深度学习优化函数详解

阅读量：501117 篇

热门文章

一文搞懂交叉熵在机器学习中的使用，透彻理解交叉熵背后的直觉

阅读量：18343

C++如何合并若干个vector

阅读量：11573

深度学习优化函数详解（3）-- mini-batch SGD 小批量随机梯度下降

阅读量：10878

深度学习优化函数详解（4）-- momentum 动量法

阅读量：10442

如何使用reCaptcha（2.0版本）来做网站验证码

阅读量：8317



高配台式电脑



联系我们



扫码联系客服

关于 招聘 广告服务 网站地图

京ICP证09002463号

百度提供站内搜索

©2018 CSDN版权所有

kefu@csdn.net

400-660-0108

QQ客服

客服论坛

经营性网站备案信息

网络110报警服务

中国互联网举报中心 北京互联网违法和不良信息举报中心

隐私政策 帮助

