从逻辑回归到受限玻尔兹曼机

原创 夕小瑶 **夕小瑶的卖萌屋** 2017-06-13

在那很久很久以前，可爱的小夕写了**一篇将逻辑回归小题大做的文章**，然后在**另一篇文章**中阐述了逻辑回归的本质，并且推广出了softmax函数。

从那之后，小夕又在**一篇文章**中阐述了逻辑回归与朴素贝叶斯的恩仇录，这两大祖先级人物将机器学习的国度划分为两大板块——**生成式与判别式**。

后来，朴素贝叶斯为了将自己的国度发扬光大，进化出了**贝叶斯网**以抗衡逻辑回归，一雪前耻。

然而，傲娇的逻辑回归怎能就此善罢甘休呢？

ps：对上面的故事有不认识的名词的同学，务必点一下上面的文章链接复习一下哦。

先复习一下逻辑回归的结论。在之前的**文章1**和**文章2**中已经解释了，逻辑回归是个二类分类器，它的假设函数是$f = sigmoid(x \cdot y)$，并且本质上这个假设函数算出来的是**其中一个类别的后验概率p(y=1|x)**。

而sigmoid函数本身并不单纯，而是一个现实意义非常丰富的函数，所以逻辑回归模型可以表示成

p(y=1|x) = exp(x与类别1的"亲密度") / exp(x与所有类别的"亲密度"之和)

其中，亲密度直接用内积x·y描述。

对上面的结论有疑问的同学，回看一下那两篇文章哦。

再次提醒！前方超级高能预警！！！
请务必在进入战场前确认已经装备以下三神器：

1、浅入深出被人看扁的逻辑回归
2、sigmoid与softmax的血缘关系
3、逻辑回归与朴素贝叶斯的战争

显然，逻辑回归这么简单的model有很大的改良余地。尤其是所谓的亲密度！

想象一下，在逻辑回归中，亲密度就是用x与y的内积来表示了，但是这个做法过于简单了。我们暂且不管最佳描述亲密度的函数是什么，我们就直接用一个函数E(x,y)来表示x与y的亲密度，然后我们尽可能的让E(x,y)的形式变得合理，尽可能的用最优的方式去描述x与y的亲密度。

首先，描述x与y的亲密度，就是描述两个向量的亲密度嘛~为了避免让大家思考的时候总是带着机器学习的影

子，我们不妨用两个一般的向量v1和v2来表示x与y。

为了找出最优的描述v1与v2亲密度的函数E(v1,v2)，我们想想以前直接用v1·v2来描述亲密度有什么缺陷。

设想一下，如果v1代表老板，v2代表老板手下的秘书呢？

显然，v1与v2的亲密度来说，v2并没有多大的发言权，老板（v1）想跟谁亲密，那么v1就跟哪个v2的亲密度大。所以！**我们需要一个权重来表示某个向量在计算亲密度时的说话分量：**

我们就用参数b来表示v1的说话分量，用参数c来表示v2的说话分量啦~

然后，再想象一下，v1和v2的亲密度完全可以体现在方方面面呀~比如老板与小王由于都喜欢美妆从而比较亲密，老板与小李都喜欢打篮球从而比较亲密，但是由于老板心里觉得美妆比篮球更重要，所以综合来看老板跟小王更亲密。

而之前用v1直接用v2做内积的话，显然向量的各个维度（篮球、美妆等各个方面）的权重都是相等的，无法描述不同维度在老板心里的权重。**那么如何分出来不同维度在老板心里的权重呢？**

显然！在v1与v2之间加个同样维度的向量描述各个维度权重！这个参数暂时用小写的w表示。诶？不对啊，如果w是个向量的话，v1、v2、w这三个向量无论怎么计算，都不可能乘出来一个表示亲密度的值啊（回想一下做矩阵乘法时的结果的维度与各乘子的维度的关系）

所以这里的w不能是向量！假如v1和v2的维度是n的话，那么**w只需要是个n*n对角矩阵**就可以啦！这样$v1^T \cdot w \cdot v2$就是维度 1*n 乘以 n*n 乘以 n*1 ，这样得到的结果就是一个值了~

所以，我们把前面的参数b和c也高级化一下，让b也能刻画不同维度下v1的分量，以及c刻画不同维度下v2的分量，所以**参数b和c就是个n维向量**啦~（$v1^T \cdot b$，就是维度1*n与维度n*1相乘，直接得到一个值。v2与c同理。）

再想想，这时参数b和c是n维向量，w是个n*n的对角矩阵。还能继续优化亲密度的描述吗？

设想一下，如果老板（v1）的第1维的含义是“喜欢化妆”，秘书（v2）的第5维的含义是“喜欢买化妆品”，那么当v1与v2直接求内积的时候，哪怕v1的第1维与v2的第5维会碰撞出强烈的亲密度，但是由于v1与v2直接求内积，也就是说v1的第i维只能跟v2的第i维碰撞，这样明显丧失了很多潜在的亲密度啊！所以我们要用额外的参数来描述v1的第i维与v2的任意的第j维之间的“关联度”，如果关联度非常大，那参数就尽可能大，让v1的第i维去尽情碰撞v2的第j维，看看能不能出来强烈的亲密度~当然，两个关联度很小的维度的话，对应的参数的值就会接近0，就没有碰撞的必要啦~碰撞的结果也没有啥影响力啦~

想的很好，那么这个复杂的参数怎么表示呢？

其实对于数学基础扎实的同学来说非常简单！参数矩阵的非对角线元素就是描述这种关系的！（有没有想起概率统计中的协方差矩阵？想起的肯定秒懂啦，没想起同学也没关系~）

所以，我们只要把**对角矩阵w变成普通的矩阵W**！这样W的对角线元素依然描述每个维度的权重，而非对角线元素就可以描述上述v1的各个维度与v2的各个维度之间的关联度啦！

通过**向量b、向量c、矩阵W**，简直是不能更完美的刻画v1与v2的亲密度了！所以综合起来，亲密度函数如下：

$$E(v1, v2) = b^T v1 + c^T v2 + v1^T \cdot W \cdot v2$$

所以，这个新的机器学习模型跟逻辑回归一样，只是把亲密度定义了一下，并且：

1、跟逻辑回归一样可以很自由的推广到多类分类的情况（不理解的同学回这篇文章复习一下sigmoid到softmax）

2、跟逻辑回归一样可以很自由的由判别式推广到生成式（不理解的同学回这篇文章复习一下逻辑回归到朴素贝叶斯）

好！那么我们就将这些进化全都用上（好疯狂...）：

1、改良亲密度的定义
2、推广到多类
3、推广到生成式

那么得到的超级模型的**假设函数**就是：

$$f = p(x = i, y = j) = \frac{e^{E(x=i,y=j)}}{\sum_{m=1}^{M} \sum_{k=1}^{K} e^{E(x=m,y=k)}}$$

其中，E(x,y)就是本文的改良版“亲密度”函数；M代表样本数量，K代表类别数量。

在这个模型中，参数就是亲密度函数中的向量b、向量c、矩阵W。

这个用尽高级技术的复杂而优美的模型叫什么呢？

这就是：**受限玻尔兹曼机（Restricted Boltzmann Machine, 即RBM）**！

其中，这里小夕讲的亲密度函数的**前面加个负号**，就是概率图模型中所谓的**能量函数**，也叫**势能函数**（理论物理中的概念），这里也是用**E(v1,v2)**表示(**新的E(v1,v2)=-旧的E(v1,v2)**)。假设函数中那个有两个求和号的恐怖大分母，就是概率图模型中的**配分函数Z**，跟小夕这里讲的意思是一模一样的，只不过用新的E(v1,v2)表示而已啦。

所以用**概率图中的表示方法，RBM的假设函数**即：

$$f = P(x = i, y = j) = \frac{1}{Z} exp(-E(x = i, y = j))$$

其中，配分函数：

$$Z = \sum_{m=1}^{M} \sum_{k=1}^{K} e^{-E(x=m,y=k)}$$

看~概率图中被包装的如此抽象的RBM，本质上就是个究极进化版的逻辑回归而已啦。

那么，为什么说这是“受限”玻尔兹曼机呢？难道还有不受限的玻尔兹曼机吗？这个答案就让小夕在下一篇文章中告诉你吧，看看RBM是如何继续进化的~

等等！

纳尼???逻辑回归培养到最后，怎么培养成了生成式模型呀！所以竟然培养成了对方战场的究级武器???

╮(￣▽￣"")╭

我想，这大概跟小夕父母的心情差不多吧——明明生的是儿子，养着养着就成女儿了（生儿育女）。