读书报告

71117203袁佳怡

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：在逻辑斯谛模型中，条件概率分布为什么选用了sigmoid函数来表示

讨论后的理解：LR通过最大化类标记的条件似然度来学习一个线性分类器，为了定义这个条件概率，使用sigmoid函数将线性分类器的响应值<w,x>映射到一个概率上。 Sigmoid函数恰好能实现这一功能；而且是一个单调上升函数，可将较大的<w,x>映射到较大的概率p(y=1|x,w)。sigmoid的这些良好性质恰好能满足LR的需求。Sigmoid函数中间上升过程是很短的一段，所以分类的样本点能够有效分类，落在相应的0和1的区间内。

1. （必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：
2. 问题2：6.2.1最大熵原理中，为什么在所有可能的概率模型（分布）中，熵最大的模型是最好的模型？

自己的理解：熵是随机变量不确定性的度量，不确定性越大，熵值就越大；若随机变量退化为定制，熵为0。均匀分布是“最不确定”的分布。两个随机变量的X，Y的联合分布，可以形成联合熵，用H(X,Y)表示，条件熵H(X|Y)=H(X,Y)-H(Y)。最大熵模型根据最大熵原理在特征限制下（约束条件）下选择最优的概率分布。最大熵模型的学习最终可以归结为以最大熵模型似然函数为目标函数的优化问题。因为此时的目标函数是凸函数，因此没有局部最优解只有全局最优解。

具体的数学推导可见：https://www.cnblogs.com/wxquare/p/5858008.html

1. 问题3：多项逻辑斯谛回归的模型公式是怎么得到的？

自己的理解：多项逻辑斯谛回归模型，是一个用来预测离散型随机变量的不同可能结果的概率的模型，用于进行多项分类。在二项逻辑斯谛回归中，分母上是1+exp的函数，而在多项的逻辑斯谛回归中，分母上是1+(k-1)个exp的函数，相当于是继承了原来的实现逻辑，但具体的数学运算目前还不是很清楚。

1. 问题4：



这里的Π（x）如何计算

自己的理解：我觉得这里的Π(x)应该只是一个概率的指代函数而已，并没有特殊的含义。

1. 问题5：



如何理解这里的fi？

自己的理解：fi应该是说，这里的模型分别要满足应该约束条件，所有的约束条件都要满足，这样的假设条件下，得到最大熵。

1. （必填）读书计划

1、本周完成的内容章节：《统计学习方法》第六章

2、下周计划：《统计学习方法》第七章

1. 读书摘要
   * 1. 逻辑斯谛回归是统计学习中的经典分类方法，而最大熵是概率模型学习的一个准则，将其推广到分类问题就得到了最大熵模型。其两者都属于对数线性模型。
     2. 逻辑斯谛分布，该分布的定义是：设X是连续随机变量，X服从逻辑斯谛分布是指X服从如下分布函数和密度函数：

* + 1. 二项逻辑斯蒂回归模型是一种分类模型，有条件概率分布P(Y|X)表示，逻辑斯谛回国模型表示如下：

其中，exp为以e为底的指数函数，x∈是输入，y∈{0，1}输出，w，b是模型参数，线性函数w·x的值越接近正无穷，概率值越接近1；反之，越接近负无穷，概率值越接近0——这就是逻辑斯谛回归模型。

* + 1. 最大熵原理：H(P)=，熵反映的是事物的无序程度，或者说随机变量分布的均匀程度。最大熵原理认为，在所有可能的概率模型中，熵最大的模型是最好的模型，也就是说，在那些满足已有事实或约束条件的概率模型中，哪些不确定的部分是无序的、混乱的、等可能的、均匀分布的，而这些需要用熵来衡量。学习的目标是用最大熵原理选择最好的分类模型。首先，模型必须满足联合分布P(X,Y)的经验分布和边缘分布P(X)的经验分布。
    2. 最大熵模型的学习：可以形式化为约束最优化问题，将最大熵模型的学习等价于约束最优化问题，等价于最小值问题，通过引入拉格朗日乘子将原始问题搞为对偶问题再极小化。
    3. 模型学习的最优化算法：常见的方法有改进的迭代尺度法、梯度下降法、牛顿法或拟牛顿法。