《统计学习方法》第6章读书笔记

一、自己提出的问题的理解：

1. 提出的问题1：逻辑斯谛回归为何采用sigmoid函数，这个函数有独特的优点吗？有没有不足之处呢？

**讨论后的理解**：逻辑斯谛函数的图形是一条S形曲线，该曲线以点为中心对称，在中心附近增长速度较快，在两端增长速度较慢。形状参数的值越小，曲线在中心附近增长得越快。

**优点**：1）逻辑斯谛回归模型目的是进行分类，因为其在切分点附近函数变化较为陡峭的性质，可提高分类的精度。

2）输入范围是，值域正好为(0,1)，满足概率分布为(0,1)的要求。

3）这是一个单调上升的函数，具有良好的连续性，不存在不连续点。

4）函数关于（0，0.5） 中心对称

**缺点**：1）幂运算相对耗时。

2）sigmoid函数反向传播时，很容易就会出现梯度消失的情况。

1. 提出的问题2：书中将最大熵模型和逻辑斯谛回归放在一起讲，两者的相似之处？

**讨论后的理解**：从最大熵的思想出发得出的最大熵模型，最后的最大化求解就是在求P(y|x)的对数似然最大化。逻辑斯谛回归也是在求条件概率分布关于样本数据的对数似然最大化。二者唯一的不同就是条件概率分布的表示形式不同。

二、别人提出的问题的理解（选择几个问题罗列，并给出理解）：

1. 问题3：逻辑斯谛算法的优点在于什么？

**个人的理解：**1）适合需要得到一个分类概率的场景。2）计算代价不高，容易理解实现。在时间和内存需求上相当高效，可应用于分布式数据，并且还有在线算法实现，用较少的资源处理大型数据。3）LR对于数据中小噪声的鲁棒性很好，并且不会受到轻微的多重共线性的特别影响。（严重的多重共线性则可以使用逻辑斯谛回归结合L2正则化来解决，但是若要得到一个简约模型，L2正则化并不是最好的选择，因为它建立的模型涵盖了全部的特征。）

1. 问题4：为何说在满足相同约束的情况下，熵最大的模型是最优的模型？

**个人的理解：**最大熵原理认为要选择的概率模型首先必须满足已有的事实，即约束条件。在没有更多信息的情况下，那些不确定的部分都是“等可能的”，即“最混乱的”，而其它的取值实质上意味着增加了其它的假设（约束）。

1. 问题5：逻辑斯谛回归模型是怎么用于多分类问题的？

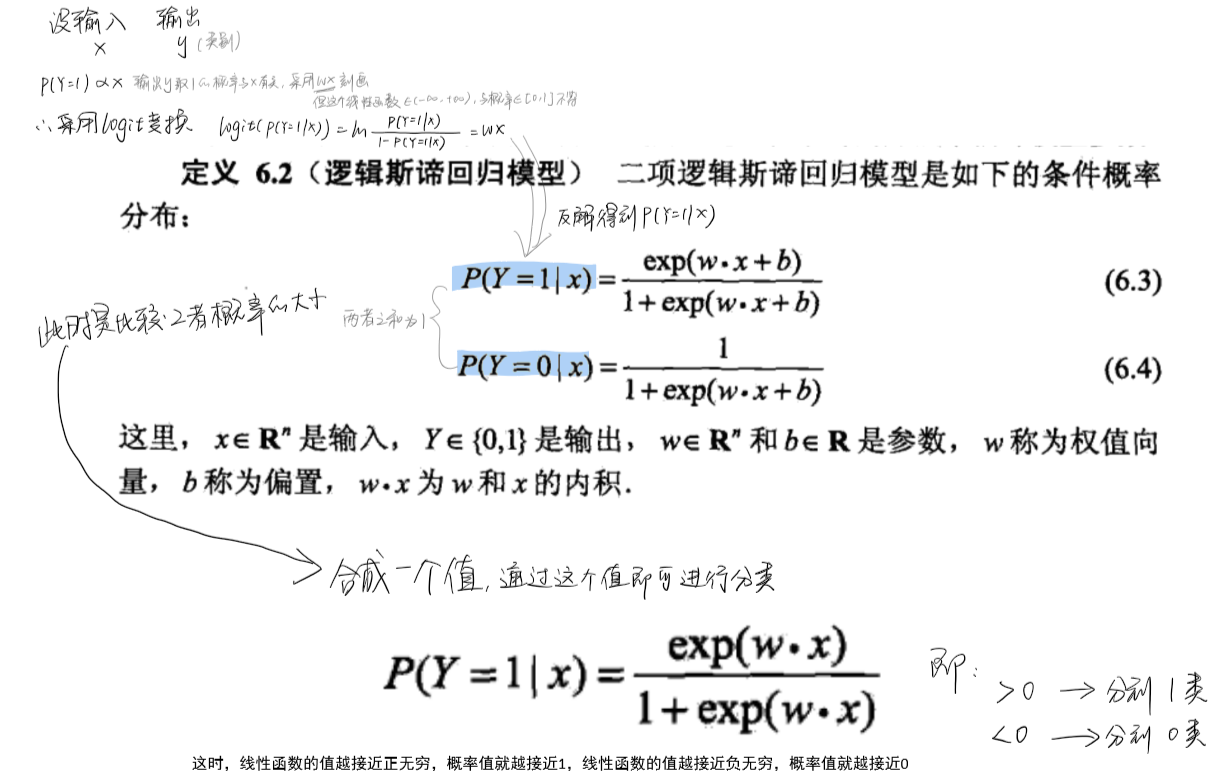
**个人的理解**：实现多类别逻辑斯谛回归模型最简单的方法是，对于所有K个可能的分类结果，我们运行K-1个独立二元逻辑斯谛回归模型，在运行过程中把其中一个类别看成是主类别，然后将其它K-1个类别和我们所选择的主类别分别进行回归。

三、读书计划

1. 本周完成的内容章节：第六章
2. 下周计划：再理解一下第六章的数学推导部分，学习第七章

四、读书摘要及理解

1）逻辑斯谛回归



2）最大熵模型

