李航统计学习：

第五章讨论部分：

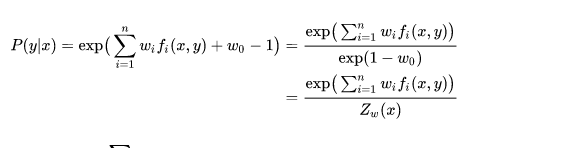
读书报告内容：

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：

第100页Zw（x）为什么等于那个式子？

讨论后的理解：

因为

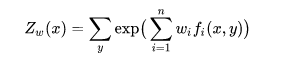


根据约束条件：



上式对y求和得到：

所以：

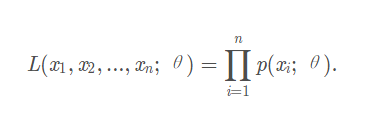


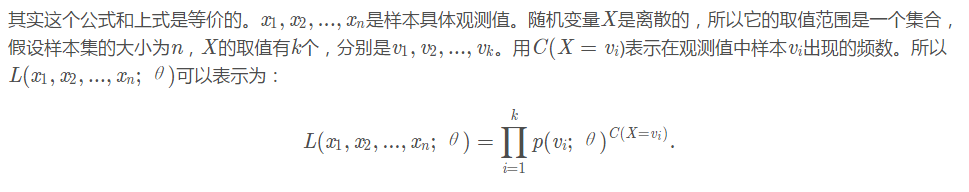
1. 提出的问题2：

102页，条件概率分布P（Y/X）的对数似然函数为什么是那个？

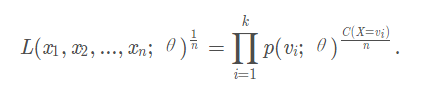
讨论后的理解：

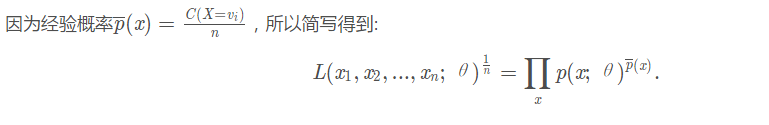
其实第一眼之所以不理解，因为这是最大似然函数的另外一种形式。一般书上描述的最大似然函数的一般形式是各个样本集 X X X中各个样本的联合概率:

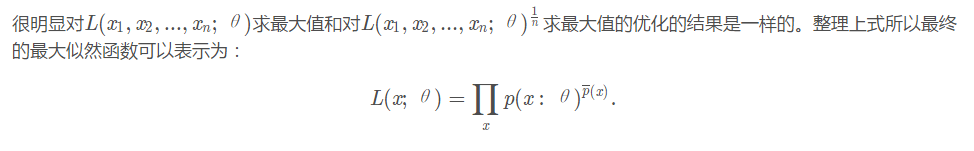




对等式两边同时开 n n n次方，可得：







二、（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）：

问题3：回归方法应用于分类有哪些难点，逻辑斯蒂回归模型是怎么解决的

自己的理解：

那么为什么不能用线性回归模型来解决分类问题呢？主要有以下三个原因：

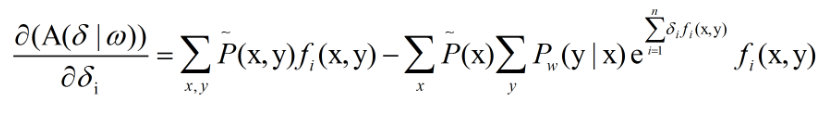
假设因变量有三个类：狗、猪、人。将狗赋值为1，猪赋值为2，人赋值为3。这种赋值方式默认了狗、猪、人是有顺序的。而且狗与猪之间的差距=猪与人之间的差距=1，这显然是有问题的。

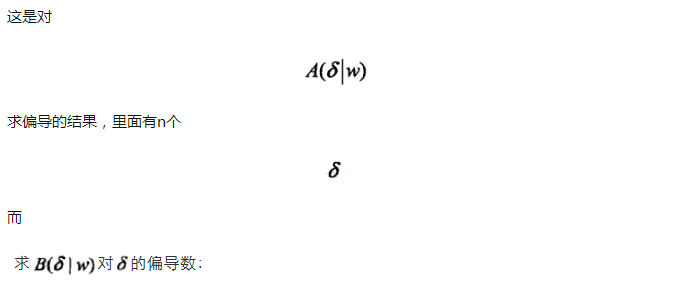
没有好的转换方法，可以将有三个及以上类的因变量转化为一个定量数据，用以进行线性回归。

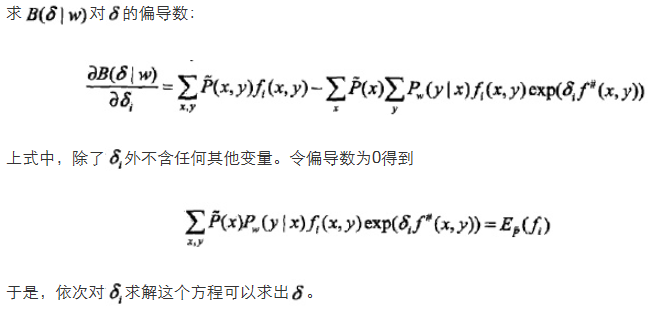
对于0/1二分类问题，线性回归得到的数值可以看作是属于这个类的概率。即如果令 [公式]，那么可以用线性模型 [公式] 进行回归，得到的 [公式] 就是 [公式] 属于类1的概率。但问题是线性回归可能会产生 [公式] 小于0或者大于1的数。

问题4：改进的迭代算法IIS主要对哪些方面进行了改进提升了效率？

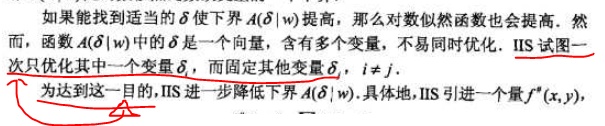
自己的理解：







而找一个更小的下界就是为了减少需要优化的变量。

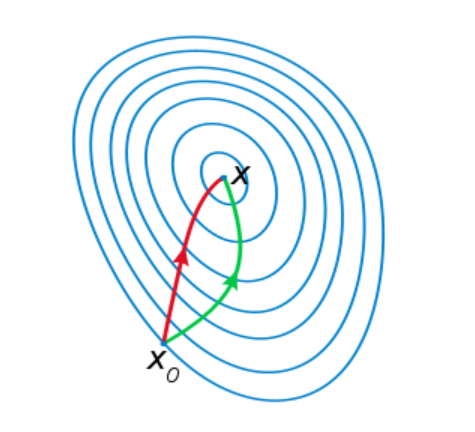


问题5：为什么牛顿和牛顿法收敛速度更快？

自己的理解：

如下图所示，红色表示牛顿法，绿色表示梯度下降法。因为梯度下降法每次只从当前所处位置选一个坡度最大的方向走一步，牛顿法在选择方向时，不仅会考虑坡度是否够大，还会考虑走了一步之后，坡度是否会变得更大。所以，可以说牛顿法比梯度下降法看得更远一点，能更快地走到最底部。

从几何上说，牛顿法就是用一个二次曲面去拟合你当前所处位置的局部曲面，而梯度下降法是用一个平面去拟合当前的局部曲面，通常情况下，二次曲面的拟合会比平面更好，所以牛顿法选择的下降路径会更符合真实的最优下降路径。



1. （必填）读书计划
2. 本周完成的内容章节：

（1）第六章完成并且梳理知识点，寻找问题，自己思考并且于小组会之前完成了自己的思考。

2、下周计划：附录ABC部分.

四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

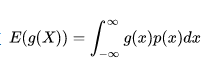
最大熵的原理：

“最大熵原理”的意思是，当我们预测一个随机变量的取值的时候，最好假设它均匀分布，保留全部不确定性，此时预测的风险最小。举个例子：如果投掷一颗骰子，让你预测各面出现的概率。在没有任何额外信息的情况下，我们会认为骰子是均匀的，各面出现的概率为1/6。从直觉上说，这是最稳妥的策略，不遗漏任何一种可能性，但如何从数学上证明这是最优策略呢？

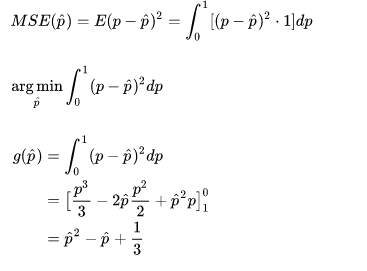
让我们来把模型简化一下，假设投掷一枚硬币，只有正反面两种结果，其中正面的概率为  ，若 p的预测值为 p ， p 应该如何取值呢？

直觉告诉我们应该取0.5，假设硬币是均匀的。为什么？

如果以真实值与预测值的“均方误差（MSE）”作为风险函数，当 p 在 [0,1] 的范围内取值时，最优策略应该使得“均方误差”的数学期望最小化：

函数的期望值

假设 p 在[0,1]范围内均匀分布，则 [公式] 的概率密度函数为 p(x)=1



显然，p=0.5时，MSE 取得最小值，即假设“硬币两面的概率相等”具有最小风险。推广至骰子，可以证明骰子“任意两面出现的概率相等”具有最小风险，即每一面概率为1/6 。