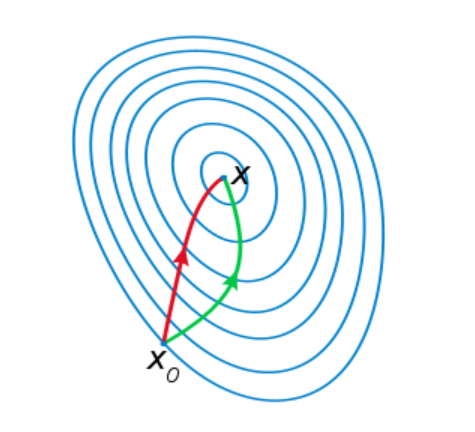
1. （必填）自己提出的问题的理解（罗列全部）：
2. **提出的问题1：**

为什么牛顿和牛顿法收敛速度更快？

讨论后的理解：

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. **提出的问题2：**

Logistic Model与Bayes Model的关系？

讨论后的理解：

逻辑回归和朴素贝叶斯的表达式十分相似，但Bayes是生成模型，Logistic是判别模型；逻辑回归和朴素贝叶斯的表达式惊人地相似，但二者在求权重时区别是很大的，朴素贝叶斯方法的条件独立假设。因为条件独立假设，朴素贝叶斯可以不使用梯度下降，而直接通过统计每个特征的逻辑发生比来作为权重。而逻辑回归，条件独立假设并不成立，通过梯度下降法，可以得到特征之间的耦合信息，从而得到相应的权重。

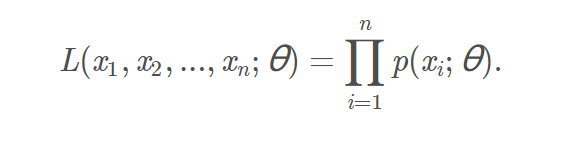
1. （必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：

1. **问题3**：

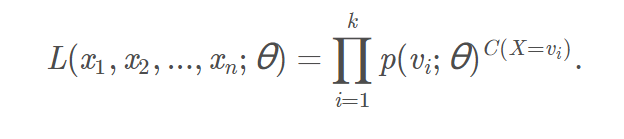
最大熵模型中的对数似然函数的解释。

自己的理解：

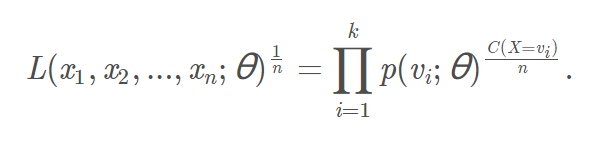
一般来说，常用的似然函数表示公式为：



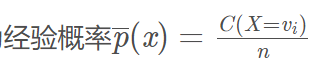
x 1 , x 2 , . . . , x n x\_1,x\_2,...,x\_nx1​,x2​,...,xn​是样本具体观测值。随机变量X是离散的，所以它的取值范围是一个集合，假设样本集的大小为n，X的取值有k个，分别是v 1 , v 2 , . . . , v k v\_1,v\_2,...,v\_kv1​,v2​,...,vk​。用C ( X = v i C(X=v\_iC(X=vi​)表示在观测值中样本v i v\_ivi​出现的频数。所以L ( x 1 , x 2 , . . . , x n ; θ ) L(x\_1,x\_2,...,x\_n;\theta)L(x1​,x2​,...,xn​;θ)可以表示为：



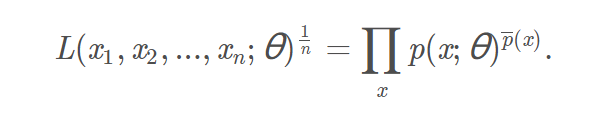
对等式两边同时开n次方，得到



因为



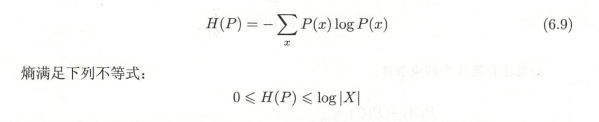
所以可以得到



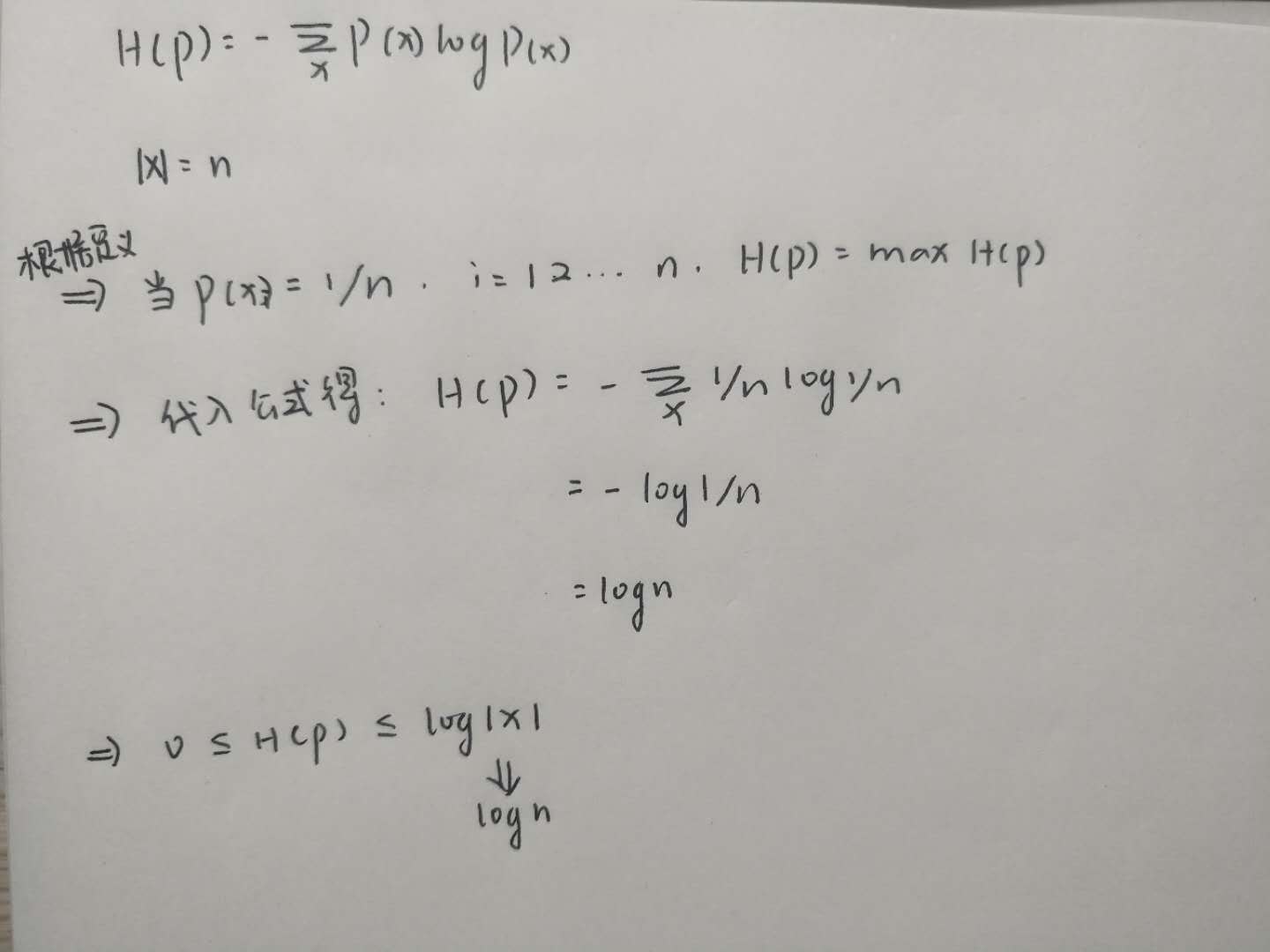
2. **问题4：**

95页式6.5下面的不等式是怎么推出来的？

自己的理解：



根据定义可以知道，熵是用来表达信息不确定性，按照这个理解，可以解释为当信息数据完全均匀分布时，不确定性最大，此时熵最大。则当分布为等概率时，即每一个样本X的概率P(i)相同都为1/n时，熵H(x)取到最大值。代入公式6.9，可以得到H(P)的最大值为log|X|。



根据定义

1. **问题5：**

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

讨论后的理解：

难点：

* 在分类模型中，对不同类别进行不同赋值相当于强制序列化，并且无法度量不同类别之间的差距。
* 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 , 那么可以用线性模型 进行回归，得到的P(X)就是Y属于类1的概率。但问题是线性回归可能会产生P(X)小于0或者大于1的数。

Logistic Model出发点是第三个问题：对于0/1二分类问题， 利用逻辑斯蒂函数（logistic function）将P(X)的范围限[0,1]之间。

三、（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：逻辑斯蒂回归与最大熵模型）

2、下周计划：统计学习方法（第七章：支持向量机）

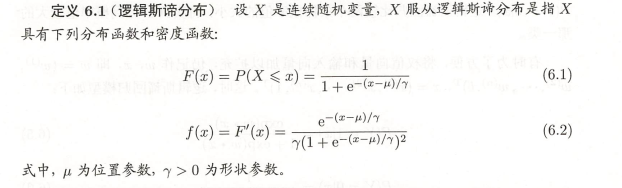
四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

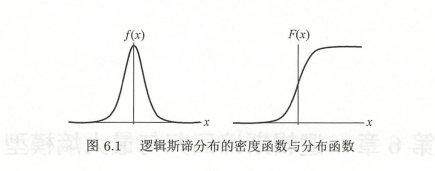
1、读书摘要及理解（选做）

本章主要介绍了逻辑斯蒂回归模型和最大熵模型

Logistic Regression 虽然被称为回归，但其实际上是分类模型，并常用于二分类。Logistic 回归的本质是：假设数据服从这个分布，然后使用极大似然估计做参数的估计。

逻辑斯蒂定义：

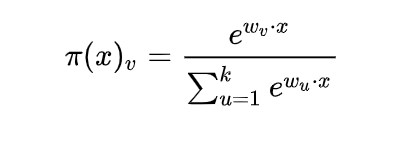




Logistic 分布是由其位置和尺度参数定义的连续分布。Logistic 分布的形状与正态分布的形状相似，但是 Logistic 分布的尾部更长，所以我们可以使用 Logistic 分布来建模比正态分布具有更长尾部和更高波峰的数据分布。在深度学习中常用到的 Sigmoid 函数就是 Logistic 的分布函数在 的特殊形式。

最大熵模型：

逻辑回归和最大熵模型本质上没有区别，最大熵在解决二分类问题时就是逻辑回归，在解决多分类问题时就是多项逻辑回归，即逻辑回归是最大熵模型的一个特殊例子。



表示，输入时 x，输出的 y=u 的概率；

A(u,v) 是一个指示函数，若 u=v，则 A(u,v)=1，否则 A(u,v)=0；

我们的目标就是从训练数据中，学习得到一个模型，使得 [公式] 最大化，也就是输入 x，预测结果是 y 的概率最大，也就是使得最大。

2、伪代码的具体实现(选做)