**一.讨论中的问题**

**1.逻辑回归的损失函数为什么要使用极大似然函数作为损失函数？**

损失函数一般有四种，平方损失函数，对数损失函数，HingeLoss0-1损失函数，绝对值损失函数。将极大似然函数取对数以后等同于对数损失函数。在逻辑回归这个模型下，对数损失函数的训练求解参数的速度是比较快的。  
因为目标是要让预测为正的的概率最大，且预测为负的概率也最大，即每一个样本预测都要得到最大的概率，将所有的样本预测后的概率进行相乘都最大，这就能到似然函数了。

**2.逻辑回归在训练的过程当中，如果有很多的特征高度相关或者说有一个特征重复了100遍，会造成怎样的影响？**

如果在损失函数最终收敛的情况下，其实就算有很多特征高度相关也不会影响分类器的效果。

**3.最大熵模型的优点有什么？**

a) 最大熵统计模型获得的是所有满足约束条件的模型中信息熵极大的模型,作为经典的分类模型时准确率较高。

b) 可以灵活地设置约束条件，通过约束条件的多少可以调节模型对未知数据的适应度和对已知数据的拟合程度

**4.逻辑斯蒂的优缺点是什么？**

1. 准确率并不是很高。因为形式非常的简单(非常类似线性模型)，很难去拟合数据的真实分布；

2. 很难处理数据不平衡的问题。举个例子：如果我们对于一个正负样本非常不平衡的问题比如正负样本比 10000:1.我们把所有样本都预测为正也能使损失函数的值比较小。但是作为一个分类器，它对正负样本的区分能力不会很好；

3. 处理非线性数据较麻烦。逻辑回归在不引入其他方法的情况下，只能处理线性可分的数据，或者进一步说，处理二分类的问题；

4. 逻辑回归本身无法筛选特征。有时候，我们会用gbdt来筛选特征，然后再上逻辑回归。

缺点：可能容易欠拟合，分类精度不高，这个可能是因为我们无法找到足够的特征

实现多类别逻辑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。

、

二、读书计划

本周第6章

下周第7章

三、摘要

本章讲了逻辑斯蒂回归以及最大熵模型

逻辑斯蒂回归：

sigmoid函数，用于隐层神经元输出，取值范围为(0,1)，它可以将一个实数映射到(0,1)的区间，可以用来做二分类。在特征相差比较复杂或是相差不是特别大时效果比较好。Sigmoid作为激活函数有以下优缺点：

优点：平滑、易于求导。

缺点：激活函数计算量大，反向传播求误差梯度时，求导涉及除法；反向传播时，很容易就会出现梯度消失的情况，从而无法完成深层网络的训练。

Logistic回归实质：发生概率除以没有发生概率再取对数。就是这个不太繁琐的变换改变了取值区间的矛盾和因变量自变量间的曲线关系。究其原因，是发生和未发生的概率成为了比值 ，这个比值就是一个缓冲，将取值范围扩大，再进行对数变换，整个因变量改变。不仅如此，这种变换往往使得因变量和自变量之间呈线性关系，这是根据大量实践而总结。所以，Logistic回归从根本上解决因变量要不是连续变量怎么办的问题。还有，Logistic应用广泛的原因是许多现实问题跟它的模型吻合。例如一件事情是否发生跟其他数值型自变量的关系。

最大熵模型：

最大熵模型的定义

最大熵原理是统计学习的一般原理,将它应用到分类就得到了最大熵模型

假设分类模型是一个条件概率分布P(Y|X),X表示输入,Y表示输出。这个模型表示的是对于给定的输入X,以条件概率P(Y|X)输出Y。

给定一个训练数据集T,我们的目标就是利用最大熵原理选择最好的分类模型。

最大熵原理指出，**对一个随机事件的概率分布进行预测时，预测应当满足全部已知的约束，而对未知的情况不要做任何主观假设。在这种情况下，概率分布最均匀，预测的风险最小，因此得到的概率分布的熵是最大**。